

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2004, 05:02:07 ; Search time 7419 Seconds

(without alignments)  
10936.536 Million cell updates/sec

Title: US-09-787-491B-18

Perfect score: 1872

Sequence: 1 gccctctgctgctgtgtgtc.....tatcaaaaaaaaaaaaaa 1872

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

```
1: GenEmbl:*
2: gb_ba:*
3: gb_hkg:*
4: gb_in:*
5: gb_om:*
6: gb_ov:*
7: gb_pat:*
8: gb_ph:*
9: gb_pl:*
10: gb_pt:*
11: gb_ro:*
12: gb_sts:*
13: gb_sy:*
14: gb_un:*
15: gb_vl:*
16: em_ba:*
17: em_fun:*
18: em_hum:*
19: em_in:*
20: em_mu:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hkg_hum:*
31: em_hkg_inv:*
32: em_hkg_other:*
33: em_hkg_mus:*
34: em_hkg_pih:*
35: em_hkg_rtd:*
36: em_hkg_mam:*
37: em_hkg_vrt:*
38: em_sy:*
39: em_hkg_hum:*
40: em_hkg_mus:*
41: em_hkg_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1872	100.0	1872	6	BD223765
2	1720	91.9	3502	6	AF419331 Human RNA
3	1483	79.2	1836	6	AX876414 Homo sapi
4	1483	79.2	1836	6	AX876414 Sequence
5	1483	79.2	1836	6	BD156116 Sequence
6	1439	76.9	1814	6	AK001286 Primer fo
7	1439	76.9	1814	6	AY007101 Homo sapi
8	1276	68.2	15745	9	AC090439 Homo sapi
9	1276	68.2	15745	9	AY048592 Homo sapi
10	1276	68.2	15745	9	AY048592 Homo sapi
11	1273	68.0	2073	9	AL590609 Human DNA
12	1225	65.4	2834	9	HSMB801468 Human DNA
13	940	50.2	1326	6	AK093226 Homo sapi
14	786	42.0	2473	6	BD223773 Homo sapi
15	786	42.0	2473	6	AX713349 Human RNA
16	726	38.8	874	9	AK054635 Homo sapi
17	674	36.0	2247	6	BC001107 Homo sapi
18	674	36.0	2247	6	AX876834 Homo sapi
19	674	36.0	2247	6	BD156341 Sequence
20	618	33.0	739	9	AK001479 Homo sapi
21	531	28.4	1633	9	AY150181 Homo sapi
22	520	27.8	1058	9	BC005039 Homo sapi
23	513	27.4	2940	9	BC010074 Homo sapi
24	509	27.2	702	6	AF419332 Homo sapi
25	509	27.2	702	6	AX868774 Sequence
26	509	27.2	2918	6	BD148836 Primer fo
27	509	27.2	2918	6	AX877218 Sequence
28	509	27.2	2918	6	BD156546 Primer fo
29	496	26.5	620	9	AK001656 Homo sapi
30	496	26.5	620	9	AF047448 Homo sapi
31	481	25.7	1556	9	AF067730 Homo sapi
32	467	24.9	789	6	AY150180 Homo sapi
33	430	23.0	737	6	AF448427 Homo sapi
34	430	23.0	737	6	BD020635 Homo sapi
35	420	22.4	881	6	BD020635 Novel gen
36	420	22.4	881	6	AX867568 Sequence
37	394	21.0	882	6	BD147630 Primer fo
38	394	21.0	882	6	AX868161 Sequence
39	370	19.8	332	6	BD148223 Primer fo
40	370	19.8	332	6	AX886243 Sequence
41	297	15.9	2562	9	BD025853 Sequence
42	297	15.9	3957	9	AK055311 Homo sapi
43	265	14.2	485	6	AK125834 Homo sapi
44	265	14.2	485	6	NR414810 Sequence
45	247	13.2	247	6	BD110363 EST and e
					AX913051 Sequence

#### ALIGNMENTS

RESULT 1  
LOCUS BD223765 1872 bp DNA linear PAT 17-JUL-2003  
DEFINITION Human RNA-associated proteins.  
ACCESSION BD223765  
VERSION BD223765.1 GI:33033535  
KEYWORDS JP 2002523045-A/1.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 1872)  
Hillman,J.L., Yue,H., Tang,Y.T., Corley,N.C., Giegler,K.J.,  
Gorgone,G.A., Patterson,C., Baughn,M.R., Ial,P., Bandman,O.,  
Reddy,R., Azimzai,Y., Shih,L.L., Yang,J. and Lu,D.A.M.



Db 1501 CCAGTATTGATCACTTATTTAAATATATGAAAAATGATTAATCTGTAAATAAATCTGG 1560  
QY 1561 TTTATTTGTCAGTGAATGATTAATACTAGATTATTAATAATTTGTTACTGCGCCACC 1620  
Db 1561 TTTATTTGTCAGTGAATGATTAATACTAGATTATTAATAATTTGTTACTGCGCCACC 1620  
QY 1621 AAACACATCTCTGATATATACCCCAAAATTAATTTAACTTTGATTTAGGTATTAAGG 1680  
Db 1621 AAACACATCTGATATATACCCCAAAATTAATTTAACTTTGATTTAGGTATTAAGG 1680  
QY 1681 AGACGTGGTGTATATTTAGTTATTTGAGGACAGAGAGCTGTATCTTAATCTGAT 1740  
Db 1681 AGACGTGGTGTATATTTAGTTATTTGAGGACAGAGAGCTGTATCTTAATCTGAT 1740  
QY 1741 TTAGTATGTTCTGTATTTAGAAAAATGTTACCAATTAATCTTTTGTAGTTATCAATG 1800  
Db 1741 TTAGTATGTTCTGTATTTAGAAAAATGTTACCAATTAATCTTTTGTAGTTATCAATG 1800  
QY 1801 TACATTTTATAGGGACATGTTCTGTATAGCGAATAATACTTTTATAGTATCAAA 1860  
Db 1801 TACATTTTATAGGGACATGTTCTGTATAGCGAATAATACTTTTATAGTATCAAA 1860  
QY 1861 AAAAAAAAAA 1872  
Db 1861 AAAAAAAAAA 1872

RESULT 2  
AF419331 3502 bp mRNA linear PRI 17-OCT-2001  
LOCUS AF419331  
DEFINITION Homo sapiens TLS-associated protein TASR-1 mRNA, complete cds,  
alternative transcript.  
ACCESSION AF419331  
VERSION AF419331.1 GI:16265856  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 3502)  
AUTHORS Clinton,J.M., Chanaky,H.A., Odell,D.D., Zielinska-Kwiatkowska,A.  
and Yang,L.  
TITLE Characterization and expression of the human gene encoding two  
JOURNAL TLS-associated serine-arginine (TASR) proteins  
REFERENCE 2 (bases 1 to 3502)  
AUTHORS Clinton,J.M., Chanaky,H.A., Odell,D.D., Zielinska-Kwiatkowska,A.  
and Yang,L.  
TITLE Direct Submission  
JOURNAL Submitted (14-SEP-2001) Department of Orthopedics, University of  
Washington, 1660 S. Columbian Way, GMR 151, Seattle, WA 98108, USA  
FEATURES  
source  
1..3502  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="NT\_004359"  
/cell\_line="Hela"  
77..628  
/note="splicing factor"  
/codon\_start=1  
/product="TLS-associated protein TASR-1"  
/protein\_id="A116665.1"  
/db\_xref="GI:16265857"  
/translation="MSRYLRPPTSIFKRVADDTSDSIDRREFGRYPIVDVYELD  
FYTRPRGPAIVQFEDVDAEDALHNDKRWICGQIEIQAGDCKTPOMAKXGR  
NVYSSSRDYDDYDRRSRSRSYERRSRSPDYNYRYSRSPNSPTGPRRSRSHS  
DNDPNCWMTQYSSAYYSRKI"  
1819..1824  
/note="alternative"  
polYA\_signal  
ORIGIN

Query Match 91.9%; Score 1720; DB 9; Length 3502;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1840; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 18 GTCTGAGGCTGCGCCCTGAGCCGCGACGGTTTGCTAGAGCCCGTTAGTGCGCCGCGC 77  
Db 1 GTCTGAGGCTGCGCCCGCTGAGCCGCGGACGGTTTGCTAGAGCCCGTTAGTGCGCCGCGC 60  
QY 78 GAGACAGCCCGCCGCGATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 137  
Db 61 GAGACAGCCCGCCGCGATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
QY 138 GAACGTGGCCGACGACACACAGTCTGAAGACTGGGGGGTGAATTTGGTCTGATGCTTC 197  
Db 121 GACGTGGCCGACGACACACAGTCTGAAGACTGGGGGGTGAATTTGGTCTGATGCTTC 180  
QY 198 TATAGTGTATGTATGTTCCACTGATTTCTACACTGCGCGTCCAGAGAGATTTGCTTA 257  
Db 181 TATAGTGTATGTATGTTCCACTGATTTCTACACTGCGCGTCCAGAGAGATTTGCTTA 240  
QY 258 TGTTCATTTTGAAGATGTTGCTGATGCTGAAGACGCTTTACATATTTTGAACAGAAAGT 317  
Db 241 TGTTCATTTTGAAGATGTTGCTGATGCTGAAGACGCTTTACATATTTTGAACAGAAAGT 300  
QY 318 GATTGTGACGCGGAGATTTGAATACAGTTTGGCCAGGGGGATGGAAGAACAACCAATCA 377  
Db 301 GATTGTGACGCGGAGATTTGAATACAGTTTGGCCAGGGGGATGGAAGAACAACCAATCA 360  
QY 378 GATGAAACCAAGGAGGAGAGATGTGTACAGTTCTTCAAGCTATGATGATATGACAG 437  
Db 361 GATGAAACCAAGGAGGAGAGATGTGTACAGTTCTTCAAGCTATGATGATATGACAG 420  
QY 438 ATACAGACGTTCTTGAAAGCCGAGTTATGAAAGAGAGATCAAGAACTCGTCTTTTGA 497  
Db 421 ATACAGACGTTCTTGAAAGCCGAGTTATGAAAGAGAGATCAAGAACTCGTCTTTTGA 480  
QY 498 TTACACTATAGAGATGCTTATGCTTACAGAAACAGTACAGCCGCTGGAACAACACGCGC 557  
Db 481 TTACACTATAGAGATGCTTATGCTTACAGAAACAGTACAGCCGCTGGAACAACACGCGC 540  
QY 558 TAG-AGAAGCATTTCCGACATGATATGACCAAACTGACGTGAAATCCCGATACGTTTC 616  
Db 541 TAGCAGAAAGCATTTCCGACATGATATGACCAAACTGACGTGAAATCCCGATACGTTTC 600  
QY 617 TGCTTACTACACTTCAAGAAAGATCTGAAGCGGAAAAAGAACCAAGAAAGGCGAGTTCA 676  
Db 601 TGCTTACTACACTTCAAGAAAGATCTGAAGCGGAAAAAGAACCAAGAAAGGCGAGTTCA 660  
QY 677 AGCGACCAAGGGGTGGAGGTGCTGAGATGTAATCTGTACGAATTTTGTGACTC 736  
Db 661 AGCGACCAAGGGGTGGAGGTGCTGAGATGTAATCTGTACGAATTTTGTGACTC 720  
QY 737 TGGTCTGAAAAAGATTAAGAAAGTGTATCGAAAACTACATGGAATTAATGAAAGTCCCTCA 796  
Db 721 TGGTCTGAAAAAGATTAAGAAAGTGTATCGAAAACTACATGGAATTAATGAAAGTCCCTCA 780  
QY 797 AGTTTGAAGTAAAGCATTTTGAACAATAATGAAGAAATTTCACTTTGTACTTGTGAAA 856  
Db 781 AGTTTGAAGTAAAGCATTTTGAACAATAATGAAGAAATTTCACTTTGTACTTGTGAAA 840  
QY 857 CTATCCCTTAAATATGATATGTTTATTTATTTGATTCATGGGTAAACAGTCCATATAATAT 916  
Db 841 CTATCCCTTAAATATGATATGTTTATTTATTTGATTCATGGGTAAACAGTCCATATAATAT 900  
QY 917 ATTGAAACTAGAGATGCTGAATATATCAAGGAACAGCCATAGTCTTCAAGTCCCTCT 976  
Db 901 ATTGAAACTAGAGATGCTGAATATATCAAGGAACAGCCATATGCTTCAAGTCCCTCT 960  
QY 977 GTTGTGCTGCTCAAACTGAATTTGGTGGAAAAAGTATGTCATATTAAGATTCAT 1036  
Db 961 GTTGTGCTGCTCAAACTGAATTTGGTGGAAAAAGTATGTCATATTAAGATTCAT 1020

QY 1037 TTTTGGCAATTTGGCAATCTTGCCCTTGTATTTTGGTCCAGTGTTCCTTAA 1096  
 DB 1021 TTTGGCAATTTGGCAATCTTGCCCTTGTATTTTGGTCCAGTGTTCCTTAA 1080  
 QY 1097 TCAATGCTTGGTGGCACTGGTATTTACTGTGACACACAGTGTTCATCTG 1156  
 DB 1081 TCAATGCTTGGTGGCACTGGTATTTACTGTGACACACAGTGTTCATCTG 1140  
 QY 1157 TCTTACTACTCTCTCCAGGTAAATTCAAATTATTTGACATCCAGCTAAGAGGCC 1216  
 DB 1141 TCTTACTACTCTCTCCAGGTAAATTCAAATTATTTGACATCCAGCTAAGAGGCC 1200  
 QY 1217 ATCTCTTCACTCTTCTTCTAGTATGATATTCGCAAAATTTTATTTAGAGCCCTTACTG 1276  
 DB 1201 ATCTCTTCACTCTTCTTCTAGTATGATATTCGCAAAATTTTATTTAGAGCCCTTACTG 1260  
 QY 1277 TGGGCAATTCATTTGATGATTAATTTGAGAAAATAGATTAATTCCTTATTCAGTAATG 1336  
 DB 1261 TGGGCAATTCATTTGATGATTAATTTGAGAAAATAGATTAATTCCTTATTCAGTAATG 1320  
 QY 1337 TCTACTGACACATCTAGTATCATTAAGATGAGCTCATTTGTTTGTGAGTGT 1396  
 DB 1321 TCTACTGACACATCTAGTATCATTAAGATGAGCTCATTTGTTTGTGAGTGT 1380  
 QY 1397 GTTATTCATTAATTTTATTAACCATTCGTATTCATTAATTAAGAACCAATATG 1456  
 DB 1381 GTTATTCATTAATTTTATTAACCATTCGTATTCATTAATTAAGAACCAATATG 1440  
 QY 1457 ATCAAGATTAAGTATTTGATGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTA 1516  
 DB 1441 ATCAAGATTAAGTATTTGATGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTA 1500  
 QY 1517 TTATTTAATTAATTAATTAATTAATTTGATGATTAATTTGATGATTAATTTGATGATTA 1576  
 DB 1501 TTATTTAATTAATTAATTAATTAATTTGATGATTAATTTGATGATTAATTTGATGATTA 1560  
 QY 1577 TGTATTAATTAATTAATTAATTTGATGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTA 1636  
 DB 1561 TGTATTAATTAATTAATTAATTTGATGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTA 1620  
 QY 1637 ATTAAGATTAAGTATTTGATGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTA 1696  
 DB 1621 ATTAAGATTAAGTATTTGATGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTA 1680  
 QY 1697 TTAGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTA 1756  
 DB 1681 TTAGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTA 1740  
 QY 1757 TTGAAGAAATTTGATGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTA 1816  
 DB 1741 TTGAAGAAATTTGATGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTA 1800  
 QY 1817 CATGCTCTGATTAAGTATTTGATGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTA 1857  
 DB 1801 CATGCTCTGATTAAGTATTTGATGATTAATTTGATGATTAATTTGATGATTAATTTGATGATTA 1841

RESULT 3  
 AX876414 1836 bp DNA linear PAT 17-DBC-2003  
 LOCUS AX876414  
 DEFINITION Sequence 11319 from Patent EP1074617.  
 ACCESSION AX876414  
 VERSION AX876414.1 GI:40031150  
 KEYWORDS  
 ORGANISM Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,  
 Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.  
 TITLE Primers for synthesizing full-length cDNA and their use  
 JOURNAL Patent: EP 1074617-A 11319 07-FEB-2001;

FEATURES  
 source  
 CDS  
 ORIGIN  
 Query Match 79.2%; Score 1483; DB 6; Length 1836;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1833; Conservative 0; Mismatches 1; Indels 3; Gaps 2;

Research Association for Biotechnology (JRP)  
 Location/Qualifiers  
 1..1836  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 73..624  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAE89263.1"  
 /db\_xref="GI:40031151"  
 /translation="MSRYLRPNTSLFVRNVAADTRESDELRFRRGRGPIVDVYPLD  
 FYTRPRGFVDFEDVDAEDALHNDKMTICRQLEIOPAGDRKTPNKKRER  
 NVSSRYDDYDRYRSRSEYRERSRFSFVNYRVSYPNRSRPTGRFRSRSHS  
 DNDPNCNMVNGYSAYTSRK1"

QY 22 GAGGCTCGGCGCGCTGAGCCGCGAGCGGTTGCTGAGGCGGTTAGTGGCCGCGCGAGA 81  
 DB 1 GAGGCTCGGCGCGCTGAGCCGCGAGCGGTTGCTGAGGCGGTTAGTGGCCGCGAGA 60  
 QY 82 CAGCGCGCGCGCATGTCCTGCTACCTGCGCTGCTCCCGCAACAGCTCTGTTGTCAGAAAC 141  
 DB 61 CAGCGCGCGCGCATGTCCTGCTACCTGCGCTGCTCCCGCAACAGCTCTGTTGTCAGAAAC 120  
 QY 142 GTGGCCGACGACACCAAGGTCGTGAAGACTTGGCGGCTGAAATTTGCTGTTATGCTTATA 201  
 DB 121 GTGGCCGACGACACCAAGGTCGTGAAGACTTGGCGGCTGAAATTTGCTGTTATGCTTATA 180  
 QY 202 GTTGAATGATATGTTCCACTGATTTTCTACACTGCGCGCGCAAGAGATTTGCTTATGTT 261  
 DB 181 GTTGAATGATATGTTCCACTGATTTTCTACACTGCGCGCGCAAGAGATTTGCTTATGTT 240  
 QY 262 CAATTTGAGAGATGTCGTGATGCTGAAGACGCTTTACATTAATTTGACAGAAAGTGGATT 321  
 DB 241 CAATTTGAGAGATGTCGTGATGCTGAAGACGCTTTACATTAATTTGACAGAAAGTGGATT 300  
 QY 322 TGTGACGCGCAGATTTGAATACAGTTTGGCCGAGGGGATCGAAAGACCAATTCAGATG 381  
 DB 301 TGTGACGCGCAGATTTGAATACAGTTTGGCCGAGGGGATCGAAAGACCAATTCAGATG 360  
 QY 382 AAAGCCAGAGAGAGAGAAATGTGACGTTTCTACGCTATGATGATTAATGACAGATAC 441  
 DB 361 AAAGCCAGAGAGAGAGAAATGTGACGTTTCTACGCTATGATGATTAATGACAGATAC 420  
 QY 442 AGACGTTTGAAGAGCGAAGTATGAAAGAGAGATCAAGAAAGTGGTCTTTGATTAAC 501  
 DB 421 AGACGTTTGAAGAGCGAAGTATGAAAGAGAGATCAAGAAAGTGGTCTTTGATTAAC 480  
 QY 502 AACTATAGAAAGATCGTATATGCTTGAAGAACGATGACCGCATCTGGAAGACACGAGTAC 560  
 DB 481 AACTATAGAAAGATCGTATATGCTTGAAGAACGATGACCGCATCTGGAAGACACGAGTAC 540  
 QY 561 AGAAGCATTCGAGCAATGATGACCAATCTGAGCTGGAATACCAAGTATGCTGCT 620  
 DB 541 AGAAGCATTCGAGCAATGATGACCAATCTGAGCTGGAATACCAAGTATGCTGCT 600  
 QY 621 TACTACACTTCAAGAAAGCTGAAAGCGGAAAGAAAGCAAGAGAGGAGTTCAGAGC 680  
 DB 601 TACTACACTTCAAGAAAGCTGAAAGCGGAAAGAAAGCAAGAGAGGAGTTCAGAGC 660  
 QY 681 ACCAAAGGAGGAGTGGTGAAGGCTGCTGATGATGATCTGATGATTAATTTGATGCTGCT 740  
 DB 661 ACCAAAGGAGGAGTGGTGAAGGCTGCTGATGATGATCTGATGATTAATTTGATGCTGCT 720  
 QY 741 CTGAAAGAGTAAAGAGTATTCGAAGAACTACATGGAATTAATTTGAGTCCCTCAAGT 800  
 DB 721 CTGAAAGAGTAAAGAGTATTCGAAGAACTACATGGAATTAATTTGAGTCCCTCAAGT 780



QY 801 TGAAGTAAAGCATTTTAAAGCAAAATAAAAGAAATTCAACTTTGTACTTGTGAAAACCTA 860  
 DB 781 TTTAAGTAAAGCATTTTAAAGCAAAATAAAAGAAATTCAACTTTGTACTTGTGAAAACCTA 840  
 QY 861 TCCCTAAATATGATAGGTTTATATGATTCAATGGGTAACAGTCCATTAATAATATG 920  
 DB 841 TCCCTAAATATGATAGGTTTATATGATTCAATGGGTAACAGTCCATTAATAATATG 900  
 QY 921 GAAATCTAGATGTCGAAATATCAAGAAAGACAGCCATAGTCTCTTACAGTGGCTGTG 980  
 DB 901 GAAATCTAGATGTCGAAATATCAAGAAAGACAGCCATAGTCTCTTACAGTGGCTGTG 960  
 QY 981 GTCGTCTCAACCTGAATTTGGTGGGAAAAGATATGTCATATATAAAGTCCATTTT 1040  
 DB 961 GTCTGTCTCAACCTGAATTTGGTGGGAAAAGATATGTCATATATAAAGTCCATTTT 1020  
 QY 1041 GCCATTAATGGCAATCTTGCTTTATTTGGTCCAGTGTCTCTGTTAAATCAT 1100  
 DB 1021 GCCATTAATGGCAATCTTGCTTTATTTGGTCCAGTGTCTCTGTTAAATCAT 1080  
 QY 1101 TTGCTTTGTGGCATCTGTGTTTATTTAATTGTACACACATGCAGTTTACATCTGCTT 1160  
 DB 1081 TTGCTTTGTGGCATCTGTG--TATTACTTGTACACACATGCAGTTTACATCTGCTT 1138  
 QY 1161 AACTATCTCTCCGAGGTAATTTCCATTAATTTGATCACTCAAGAGGCCCACATCT 1220  
 DB 1139 AACTATCTCTCTCCGAGGTAATTTCCATTAATTTGATCACTCAAGAGGCCCACATCT 1198  
 QY 1221 CTTCCTCACTCTTCTAGTCAGTATATTCAGCAAAATTTATTTAGAGCCCTTACCTGAGG 1280  
 DB 1199 CTTCCTCACTCTTCTAGTCAGTATATTCAGCAAAATTTATTTAGAGCCCTTACCTGAGG 1258  
 QY 1281 CAAATCATTTGTCGATTAATTTGAGAAAATATGATTAATCCCTTATTCAGTAATATGCTA 1340  
 DB 1259 CAAATCATTTGTCGATTAATTTGAGAAAATATGATTAATCCCTTATTCAGTAATATGCTA 1318  
 QY 1341 CTGAGCACAATCTAGTAATCATATACAGTATAGGCTCATTTGTTTGAAGTGTGTTA 1400  
 DB 1319 CTGAGCACAATCTAGTAATCATATACAGTATAGGCTCATTTGTTTGAAGTGTGTTA 1378  
 QY 1401 TTCAATCAATATTTTACACCATTCGTATCAATATTAATTTAGAACCAATATACGATCA 1460  
 DB 1379 TTCAATCAATATTTTACACCATTCGTATCAATATTAATTTAGAACCAATATACGATCA 1438  
 QY 1461 AGGATTAAGTATTTGTGTGTTATCTGCATTTAAAGATCAAGTATTTGATCAATTAAT 1520  
 DB 1439 AGGATTAAGTATTTGTGTGTTATCTGCATTTAAAGATCAAGTATTTGATCAATTAAT 1498  
 QY 1521 TATTAATTAATGAAAAATGATTTAATCTGTATTAATCTGTTTATTTGTGACGTGTA 1580  
 DB 1499 TATTAATTAATGAAAAATGATTTAATCTGTATTAATCTGTTTATTTGTGACGTGTA 1558  
 QY 1581 ATATACTAGATTAATTAATTTGTTTACTGTGCTTCAACCAACACATGCTAGGATTA 1640  
 DB 1559 ATATACTAGATTAATTAATTTGTTTACTGTGCTTCAACCAACACATGCTAGGATTA 1618  
 QY 1641 CCCCCAAAATTAATTTTAACTTTGATTAAGTATTAAGAGAGCTGGTCTTAATTAATG 1700  
 DB 1619 CCCCCAAAATTAATTTTAACTTTGATTAAGTATTAAGAGAGCTGGTCTTAATTAATG 1678  
 QY 1701 ATATTTTGAAGAGACAGAGAGCTGTATTCCTAATCGATTTAGTATGTTCTGTAATGA 1760  
 DB 1679 ATATTTTGAAGAGACAGAGAGCTGTATTCCTAATCGATTTAGTATGTTCTGTAATGA 1738  
 QY 1761 GAAATATTTCAACCAATTTACTTTTATGATTTACATGATCACTTTTATAGGAGACAG 1820  
 DB 1739 GAAATATTTCAACCAATTTACTTTTATGATTTACATGATCACTTTTATAGGAGACAG 1798  
 QY 1821 TTTCTGTATAGGCAATTAATACTTTTATAGTATCA 1857  
 DB 1799 TTTCTGTATAGGCAATTAATACTTTTATAGTATCA 1835

RESULT 4  
 BD156116  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,  
 Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.  
 Primer for synthesizing full-length cDNA and use thereof  
 Patent: JP 2002191363-A 10959 09-JUL-2002;  
 HELIX RESEARCH INSTITUTE  
 OS Homo sapiens (human)  
 PN JP 2002191363-A/10959  
 PD 09-JUL-2002  
 PF 28-JUL-2000 JP 2000280990  
 PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU  
 PI SAITO,  
 PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,  
 PI KEIICHI NAGAI,TETSUJI OTSUKI  
 PC

C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC  
 10',C12P21/02,C1201/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC  
 Primer for synthesizing full-length cDNA and use thereof FH Key  
 Location/Qualifiers  
 FT CDS (73)..(621).  
 source  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

## ORIGIN

Query Match 79.2%; Score 1483; DB 6; Length 1836;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1833; Conservative 0; Mismatches 1; Indels 3; Gaps 2;

QY 22 GAGGCTCGGCGGCTGAGCGGCGGAGCGTTTGTGAGCCCGTTATAGTCCGCCGCGAGA 81  
 DB 1 GAGGCTCGGCGGCTGAGCGGCGGAGCGTTTGTGAGCCCGTTATAGTCCGCCGCGAGA 60  
 QY 82 CACGCCGCGCATGTCCGCTACCTGCGTCCCCCAACACAGCTCTGTTGTCAGGAGC 141  
 DB 61 CACGCCGCGCATGTCCGCTACCTGCGTCCCCCAACACAGCTCTGTTGTCAGGAGC 120  
 QY 142 GTGGCCGAGACACACAGTGTGAAGACTGCGGCGTGAATTTGTCGTTATAGTCTATA 201  
 DB 121 GTGGCCGAGACACACAGTGTGAAGACTGCGGCGTGAATTTGTCGTTATAGTCTATA 180  
 QY 202 GTTGATGTATATTTCACTGATTTCTACACTGCGCGCTCCAGAGATTTGCTATAGTT 261  
 DB 181 GTTGATGTATATTTCACTGATTTCTACACTGCGCGCTCCAGAGATTTGCTATAGTT 240  
 QY 262 CAATTTGAGGATGTTGATGATGCTGAAGACGCTTTACATATTTTGACAGAAAGTGATT 321  
 DB 241 CAATTTGAGGATGTTGATGATGCTGAAGACGCTTTACATATTTTGACAGAAAGTGATT 300  
 QY 322 TGTGACGCGCAATTTGAATTAAGTTTCCAGGCGGAGATCGAAGACACCAATCGATG 381  
 DB 301 TGTGACGCGCAATTTGAATTAAGTTTCCAGGCGGAGATCGAAGACACCAATCGATG 360  
 QY 382 AAAGCCAAAGGAGGAGGAGATGATCAAGTCTTCAACGCTATAGCATTAATAGACAGATAC 441  
 DB 361 AAAGCCAAAGGAGGAGGAGATGATCAAGTCTTCAACGCTATAGCATTAATAGACAGATAC 420  
 QY 442 AAGCGTTTGAAGCCGAAATTATGAAAGAGAGATCAAGAGTGGCTTTGATTAC 501

Db 421 AGACGTTCTAGAAAGCCGAACTTATGAAAAGAGAGATCAGAGAGTCCGCTCTTTGATTAC 480  
 QY 502 AACTATAGAGATCGTAAAGCTCTAGAAAACAGTAGACGCACTGGAGACCAAGCGCTAG- 560  
 Db 481 AACTATAGAGATCGTAAAGCTCTAGAAAACAGTAGACGCACTGGAGACCAAGCGCTAG- 540  
 QY 561 AGAAGCAATCCGACATGATGACCCAACTGCGAGCTGGAAATACCAGTACAGTTCTGCT 620  
 Db 541 AGAAGCAATCCGACATGATGACCCAACTGCGAGCTGGAAATACCAGTACAGTTCTGCT 600  
 QY 621 TACTACACTTCAAAAAGATCTGAAAGCGGAAAAAGAAACCAAGAGGGGAGTTCAAGCG 680  
 Db 601 TACTACACTTCAAAAAGATCTGAAAGCGGAAAAAGAAACCAAGAGGGGAGTTCAAGCG 660  
 QY 681 ACCAAAGGGTGGGAGAGGCTGCTGCACTATGAACTACTGACGATATTTTGAATTTTGA 740  
 Db 661 ACCAAAGGGTGGGAGAGGCTGCTGCACTATGAACTACTGACGATATTTTGAATTTTGA 720  
 QY 741 CTGAAAAATAAAGATGTTATGCAAAAATCACTGAAATTAATTTGAATTTTGAATTTTGA 800  
 Db 721 CTGAAAAATAAAGATGTTATGCAAAAATCACTGAAATTAATTTGAATTTTGAATTTTGA 780  
 QY 801 TGAAGATGAGATTTTGAAGCAAAATTAAGGAAATTTCACTTGTACTTGTGAAACTAA 860  
 Db 781 TGAAGATGAGATTTTGAAGCAAAATTAAGGAAATTTCACTTGTACTTGTGAAACTAA 840  
 QY 861 TCCCTAAATATGATAGGTTTATATTTGATTCATGGGTAACAGGTCATATTAATTTATG 920  
 Db 841 TCCCTAAATATGATAGGTTTATATTTGATTCATGGGTAACAGGTCATATTAATTTATG 900  
 QY 921 GAAATCTAGAGATCTGGAATATATCAAGAAACAGCCATGCTCTTACAGTGGCTGCTGTT 980  
 Db 901 GAAATCTAGAGATCTGGAATATATCAAGAAACAGCCATGCTCTTACAGTGGCTGCTGTT 960  
 QY 981 GCTGCTTCAAACTGAAATGAGTGGGAAAAAGTATGTCATATTAATTAAGTTCCATTTT 1040  
 Db 961 GCTGCTTCAAACTGAAATGAGTGGGAAAAAGTATGTCATATTAATTAAGTTCCATTTT 1020  
 QY 1041 GCCATTATTTGGCAAACTTTGCTTTGTTATTTTGGTGCAGTGTCTTCTCTTAATCAT 1100  
 Db 1021 GCCATTATTTGGCAAACTTTGCTTTGTTATTTTGGTGCAGTGTCTTCTCTTAATCAT 1080  
 QY 1101 TTGCTTTGTTGGCAGTCTGCTTTATTTATTTGATACCAAGCAAGCAAGTTCATCTGCT 1160  
 Db 1081 TTGCTTTGTTGGCAGTCTGCTTTATTTATTTGATACCAAGCAAGCAAGTTCATCTGCT 1138  
 QY 1161 AACTACTCTCTCCAGGTAATTTCCAAATTAATTTGACATCCAGCTAAGAGGCGCCATCT 1220  
 Db 1139 AACTACTCTCTCCAGGTAATTTCCAAATTAATTTGACATCCAGCTAAGAGGCGCCATCT 1198  
 QY 1221 CTTTCTACCTCTTTCTAGTCAATATTTGCAAAATTTTATTTGAGCCCTTACTGTGGG 1280  
 Db 1199 CTTTCTACCTCTTTCTAGTCAATATTTGCAAAATTTTATTTGAGCCCTTACTGTGGG 1258  
 QY 1281 CAATATATGTCATGATATTTGAGAAAAATGATATTTCCCTTATTCAGTAAATGTCTA 1340  
 Db 1259 CAATATATGTCATGATATTTGAGAAAAATGATATTTCCCTTATTCAGTAAATGTCTA 1318  
 QY 1341 CTGAGACCAATCTAGGAATCATTTACAGATAGGCTCATTTGTTTGGAGGTGTGTTA 1400  
 Db 1319 CTGAGACCAATCTAGGAATCATTTACAGATAGGCTCATTTGTTTGGAGGTGTGTTA 1378  
 QY 1401 TTGATACCAATATTTTACACCAATTTGATATTTTATTTAACAACAATTAACATCA 1460  
 Db 1379 TTGATACCAATATTTTACACCAATTTGATATTTTATTTAACAACAATTAACATCA 1438  
 QY 1461 AGATATAGTATTTGAGGTTATCTGCAATTTTAAAGTATCCAGTATTTGATCACTAT 1520  
 Db 1439 AGATATAGTATTTGAGGTTATCTGCAATTTTAAAGTATCCAGTATTTGATCACTAT 1498  
 QY 1521 TATTAATATGAAAAATGATTTATCTGTAATTAACCTGGTTTATTTGTCAGTGA 1580  
 Db 1499 TATTAATATGAAAAATGATTTATCTGTAATTAACCTGGTTTATTTGTCAGTGA 1558

QY 1581 ATATCTAGAGTATTAATTAATTTGTTTACCTGCTCACCACAAACAGATGCTAGATATTA 1640  
 Db 1559 ATATCTAGAGTATTAATTAATTTGTTTACCTGCTCACCACAAACAGATGCTAGATATTA 1618  
 QY 1641 CCCCCAAATTAAGTATTTAATTTGATTTGATTTAGTATTAAGAGAGTGGTGTCTATTAATTA 1700  
 Db 1619 CCCCCAAATTAAGTATTTAATTTGATTTGATTTAGTATTAAGAGAGTGGTGTCTATTAATTA 1678  
 QY 1701 ATATTTTGAAGGACAGAGAGAGTGTATCTTAACTGATTTAGTATTTAGTATTTAGTATTTGA 1760  
 Db 1679 ATATTTTGAAGGACAGAGAGAGTGTATCTTAACTGATTTAGTATTTAGTATTTAGTATTTGA 1738  
 QY 1761 GAAATGTTACCAAAATTAATTTAATTTAGTATTTAATTTAATTTAATTTAATTTAATTTA 1820  
 Db 1739 GAAATGTTACCAAAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 1798  
 QY 1821 TTCTGTATAGCGAATTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 1857  
 Db 1799 TTCTGTATAGCGAATTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 1835

RESULT 5  
 AK001286  
 LOCUS  
 DEFINITION Homo sapiens TLS-associated protein TARS-2 mRNA.  
 ACCESSION AK001286  
 VERSION AK001286.1 GI:7022445  
 KEYWORDS oligo cloning; fls (full insert sequence).  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS  
 1 Isogai, T., Oka, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H., Wadatsuna, M., Hosokiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K., Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahata, K., Masuno, Y., Ninomiya, K. and Iwayanagi, T.  
 NEBO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 1836)  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 JOURNAL  
 TITLE  
 Isogai, T. and Otsuki, T.  
 Direct Submission  
 Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
 COMMENT  
 NEBO human cDNA sequencing project supported by Ministry of International Trade and Industry for Biotechnology; cDNA library sequencing; Research Association for Biotechnology; cDNA library construction; 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.  
 FEATURES  
 Location/Qualifiers  
 1..1836  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="NT2RP1000272"  
 /cell\_line="NT2"  
 /cell\_type="teratocarcinoma"  
 /clone\_11b="NT2RP1"  
 /note="Cloning vector: pUC19/PL3-mRNA from NT2 neuronal precursor cells after 48-hours retinoic acid (RA) induction."  
 73..624  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAA91601.1"

CDS

ORIGIN

/db\_xref="GI:7022446"  
/translation="MSRYLRPPTSLFVRNVAADTRSEDLRREFRGRIVDVVPLD  
FYTRPRGFAVQFEDVRDAEDALHLNRKMLCGRIEIQFQGRKTPNOKAKRGR  
NYSRRDDYDRYRSTRSRYSRRSRSRSPDYRYSRSPRNSRPTGRPRSRSHS  
DNDRBNCWNTQYSSAYYSRKI"

Query Match 79.2%; Score 1483; DB 9; Length 1836;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1833; Conservative 0; Mismatches 1; Indels 3; Gaps 2;

QY 22 GAGGCTCGGCGCGCTGAGCGCGGAGCGGTTGCTGAGCCCGTTAGTGGCCCGCGGAG 81  
Db 1 GAGGCTCGGCGCGCTGAGCGCGGAGCGGTTGCTGAGCCCGTTAGTGGCCCGCGGAG 60

QY 82 CAGCGCGCGCGCATGTCGCGTACCTGCGTCCCGCCCAACAGTCTCTGTGTGTCAGAAC 141  
Db 61 CAGCGCGCGCGCATGTCGCGTACCTGCGTCCCGCCCAACAGTCTCTGTGTGTCAGAAC 120

QY 142 GTGGCGCGAGACACAGGCTCTGAAGAATTGGCGCGCTGAATTTGCTTATGCTCTATA 201  
Db 121 GTGGCGCGAGACACAGGCTCTGAAGAATTGGCGCGCTGAATTTGCTTATGCTCTATA 180

QY 202 GTTGAATGTATGTTCCACTGATTTCTACACTGCGCGCTCCAGAGGATTTGCTTATGTT 261  
Db 181 GTTGAATGTATGTTCCACTGATTTCTACACTGCGCGCTCCAGAGGATTTGCTTATGTT 240

QY 262 CAATTTGAGAGTGTCTGATGCTGAGACGCTTACATTAATTTGAGACAGAAAGTGAAT 321  
Db 241 CAATTTGAGAGTGTCTGATGCTGAGACGCTTACATTAATTTGAGACAGAAAGTGAAT 300

QY 322 TGTGACGCGCGAGTTGAAATACAGTTTGCCAGAGGCGGATCGAAAGACCAAAATCAGATG 381  
Db 301 TGTGACGCGCGAGTTGAAATACAGTTTGCCAGAGGCGGATCGAAAGACCAAAATCAGATG 360

QY 382 AAAGCGAAGAGGAGGAGATGTGACAGTCTTCAAGCTATGATGATTAAGACAGATAC 441  
Db 361 AAAGCGAAGAGGAGGAGATGTGACAGTCTTCAAGCTATGATGATTAAGACAGATAC 420

QY 442 AGACGTTCTAGAACCGGAGTTTAAAGAGAGAGATCAAGAGTGGTCTTTTGAATAC 501  
Db 421 AGACGTTCTAGAACCGGAGTTTAAAGAGAGAGATCAAGAGTGGTCTTTTGAATAC 480

QY 502 AACTATGAGAGATGTATGCTGTAAGAAAGATGACGAGTGAAGACCAAGCGGATG- 560  
Db 481 AACTATGAGAGATGTATGCTGTAAGAAAGATGACGAGTGAAGACCAAGCGGATG- 540

QY 561 AGAGCCATTCGACATGATGAGCAAACTGCGAGCTGGAATACCCAGTACAGTTCGCT 620  
Db 541 AGAGCCATTCGACATGATGAGCAAACTGCGAGCTGGAATACCCAGTACAGTTCGCT 600

QY 621 TACTACATTCAGAAAGATCTGAAAGCGGAAAGAAACAAAGAGGCGAGTCAAGCG 680  
Db 601 TACTACATTCAGAAAGATCTGAAAGCGGAAAGAAACAAAGAGGCGAGTCAAGCG 660

QY 681 ACCAAAGGCGGTGGAAGGCTGCTGAGTGAATGAATCTGTAAGAAATTTTGACTGCT 740  
Db 661 ACCAAAGGCGGTGGAAGGCTGCTGAGTGAATGAATCTGTAAGAAATTTTGACTGCT 720

QY 741 CTGAAGAGTAAAGATGTTATGAAAGCTGAGATGAATGAATGAAGTCCCTCAAGT 800  
Db 721 CTGAAGAGTAAAGATGTTATGAAAGCTGAGATGAATGAATGAAGTCCCTCAAGT 780

QY 801 TGAAGTAAAGATTTTGAAGCAATTAAGAAATTTCACTTGTACTGTGGAAGCTAA 860  
Db 781 TGAAGTAAAGATTTTGAAGCAATTAAGAAATTTCACTTGTACTGTGGAAGCTAA 840

QY 861 TCCCTAAATTAAGATGTTATGATGATGATGATGATGATGATGATGATGATGATGAT 920  
Db 841 TCCCTAAATTAAGATGTTATGATGATGATGATGATGATGATGATGATGATGATGAT 900

QY 921 GAAACTGAGATGCTGAATATCAAGGAAGACAGCAATGCTCTTAAAGGCTCTGCTG 980

Db 901 GAAACTGAGATGCTGAATATCAAGGAAGACAGCAATGCTCTTAAAGGCTCTGCTG 960

QY 981 GCTGCTCAACCTGAATTTGGGTGGGAAAGAGTATGCTCAATATTAAGATTCATTTT 1040  
Db 961 GCTGCTCAACCTGAATTTGGGTGGGAAAGAGTATGCTCAATATTAAGATTCATTTT 1020

QY 1041 GCCATTAATGGCAATCTTGCCTTGTATTTTGGGCGAGTGTCTTCTGCTTATCAT 1100  
Db 1021 GCCATTAATGGCAATCTTGCCTTGTATTTTGGGCGAGTGTCTTCTGCTTATCAT 1080

QY 1101 TTGCTTGTGGGATCTGTGTTATTTATCTGTACACCAATGACAGTATTAATCTGTCT 1160  
Db 1081 TTGCTTGTGGGATCTGTG--TATTTACTGTACACCAATGACAGTATTAATCTGTCT 1138

QY 1161 AACTATCTCTTCCAGGATTAATTTCAATTAATTTGACATCCAGCTTAAGAGGCCCATCT 1220  
Db 1139 AACTATCTCTTCCAGGATTAATTTCAATTAATTTGACATCCAGCTTAAGAGGCCCATCT 1198

QY 1221 CTTCTCACTCTTCTCTGAGTATTTCAAGCAAAATTTTATGAGGCCCTTACGTGGG 1280  
Db 1199 CTTCTCACTCTTCTCTGAGTATTTCAAGCAAAATTTTATGAGGCCCTTACGTGGG 1258

QY 1281 CAATTCATTTGATGATTAATTAAGAAATATGATTAATTTCCCTTATTCAGTAATGCTCA 1340  
Db 1259 CAATTCATTTGATGATTAATTAAGAAATATGATTAATTTCCCTTATTCAGTAATGCTCA 1318

QY 1341 CTGAGCAATCTCTGATTAATTAAGATTAAGGCTCATTTGTTTGTGAGGTGTGTTA 1400  
Db 1319 CTGAGCAATCTCTGATTAATTAAGATTAAGGCTCATTTGTTTGTGAGGTGTGTTA 1378

QY 1401 TTGATTAACAATATTTTACACCATTCGATCAATGATTAATTAAGACCAATATACATCA 1460  
Db 1379 TTGATTAACAATATTTTACACCATTCGATCAATGATTAATTAAGACCAATATACATCA 1438

QY 1461 AGATTAAGATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTA 1520  
Db 1439 AGATTAAGATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTA 1498

QY 1521 TATTAATTAAGAAATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTAATTTG 1580  
Db 1499 TATTAATTAAGAAATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTAATTTG 1558

QY 1581 ATATACAGATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTA 1640  
Db 1559 ATATACAGATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTA 1618

QY 1641 CCCCCAAATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTA 1700  
Db 1619 CCCCCAAATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTA 1678

QY 1701 ATTAATTTGAGGAGACAGAGAGCTGTATCTTACCTGATTTAGTATGTTCTGTAATTTGA 1760  
Db 1679 ATTAATTTGAGGAGACAGAGAGCTGTATCTTACCTGATTTAGTATGTTCTGTAATTTGA 1738

QY 1761 GAAATATGTTACCAATTTTACTTTTATGATTTTACATGATTAATTTTATTAAGGGAATG 1820  
Db 1739 GAAATATGTTACCAATTTTACTTTTATGATTTTACATGATTAATTTTATTAAGGGAATG 1798

QY 1821 TTCTGTGTATGAGGATTAATTAATTTTATTAATTTGATTAATTTGATTAATTTGATTA 1857  
Db 1799 TTCTGTGTATGAGGATTAATTAATTTTATTAATTTGATTAATTTGATTAATTTGATTA 1835

# RESULT 6

AY007101 1814 bp mRNA linear PRI 31-AUG-2000  
DEFINITION Homo sapiens clone TCCCIA00269 mRNA sequence.  
ACCESSION AY007101  
VERSION AY007101.1 GI:9955992  
KEYWORDS FLI CDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



```

QY 1612 TGCCTCACCAACACATGCTAGGATATTAACCCCAAAATTAAGTATTAACTTGCATTG 1671
DB 1558 TGCCTCACCAACACATGCTAGGATATTAACCCCAAAATTAAGTATTAACTTGCATTG 1617
QY 1672 GTATTAAGAGAGACTGGCTGCTATTAATTAGATTATTTTGGAGACAGAGAGCTTTATC 1731
DB 1618 GTATTAAGAGAGACTGGCTGCTATTAATTAGATTATTTTGGAGAGAGAGAGCTTTATC 1677
QY 1732 CTAATCTGATTGATGATGTTCTGTATTAATGAGAAAATGTTACCAAAATTAATCTTTAGTG 1791
DB 1678 CTAATCTGATTGATGATGTTCTGTATTAATGAGAAAATGTTACCAAAATTAATCTTTAGTG 1737
QY 1792 ATTACAGTATGATCTTTTATTAAGGGGACATGTTCTGTATTAAGGATTAATACTTTATA 1851
DB 1738 ATTACAGTATGATCTTTTATTAAGGGGACATGTTCTGTATTAAGGATTAATACTTTATA 1797
QY 1852 GTA 1854
DB 1798 GTA 1800

RESULT 7
AC090439/c 199002 bp DNA linear HTG-29-JUN-2002
LOCUS Homo sapiens chromosome 12 clone RP11-418117, WORKING DRAFT
DEFINITION AC090439.8 GI:21629092
ACCESSION AC090439
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 199002)
Muzny,D.M., Adams,C., Adio-Obiola,B., Ali-osman,F.R., Allen,C.,
Alshrocks,S.L., Amaralunga,H.C., Are,J.R., Ayale,M., Banks,F.,
Barbata,J., Benton,J., Bimaga,K., Blakenburg,K., Bonnin,D.,
Bouch,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denu,A.L., Ding,Y., Dinh,H.H.,
Doutwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hult,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karleson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,Z., Lichtenage,O., Lien,C., Liu,J., Liu,W., Louleaged,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,K., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogul,M., Okwundu,G.,
Ogunye,N., Oyiedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peterson,L., Pickens,R., Primus,E., Pu,J.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojupokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,
Sodergren,E., Somaite,T., Sparks,A., Stanley,H., Stone,I.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,O.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wlaczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 199002)
Worley,K.C.
Direct Submission
Submitted (22-FEB-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 199002)
Worley,K.C.
Direct Submission
Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 29, 2002 this sequence version replaced gi:20279190.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Drafting Center Code: WIR
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HDAC
Center clone name: RP11-418117
----- Summary Statistics
Sequencing vector: M13;
Chemistry: Dye-Primer Body: 65% of reads
Chemistry: Dye-terminator Big Dye: 35% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 195705 bases at least Q40
Consensus quality: 197756 bases at least Q30
Consensus quality: 199521 bases at least Q20
Estimated insert size: 211863; sum-of-contigs estimation
Quality coverage: 9x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2051: contig of 2051 bp in length
2052 2151: gap of unknown length
2152 4816: contig of 2665 bp in length
4817 4916: gap of unknown length
4917 8256: contig of 3340 bp in length
8257 8356: gap of unknown length
8357 11062: contig of 2706 bp in length
11063 11162: gap of unknown length
11163 14224: contig of 3062 bp in length
14225 14324: gap of unknown length
14325 18185: contig of 3861 bp in length
18186 18285: gap of unknown length
18286 23102: contig of 4817 bp in length
23103 28667: gap of unknown length
28668 28767: contig of 5465 bp in length
28768 34634: gap of unknown length
34635 34734: gap of unknown length
34735 40230: contig of 5456 bp in length
40231 40330: gap of unknown length
40331 46114: contig of 5784 bp in length
46115 46214: gap of unknown length
46215 53091: contig of 6877 bp in length
53092 53191: gap of unknown length
53192 125180: contig of 71989 bp in length
125181 199002: gap of unknown length
199002: contig of 73722 bp in length.

```

FEATURES  
Source Location/Qualifiers  
1..19902  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-418117"

## ORIGIN

Query Match 76.9%; Score 1439; DB 2; Length 19902;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1799; Conservative 0; Mismatches 0; Indels 4; Gaps 2;

QY 53 GCTGAGCCCGTTAGTGGCCGCGCGAGACACGCGCCGCAATGCCGCTACTCGTCG 112  
Db 22174 GCTGAGCCCGTTAGTGGCCGCGCGAGACACGCGCCGCAATGCCGCTACTCGTCG 22115  
QY 113 CCCCCACACGCTCTGTTGCTGCAAGAACGTCGCGACGACACAGGTCTGAAAGCTTGC 172  
Db 22114 CCCCCACACGCTCTGTTGCTGCAAGAACGTCGCGACGACACAGGTCTGAAAGCTTGC 22055  
QY 173 GGCCTGAATTTGTCGTATGAGTCCCTATAGTGTGATGTGTCACCTGATTTCTACA 232  
Db 22054 GGCCTGAATTTGTCGTATGAGTCCCTATAGTGTGATGTGTCACCTGATTTCTACA 21995  
QY 233 CTCGCGCTCCAAAGAGATTGCTTATGTTCAATTTGAGAGTGTGCTGATGCTGAAGCG 292  
Db 21994 CTCGCGCTCCAAAGAGATTGCTTATGTTCAATTTGAGAGTGTGCTGATGCTGAAGCG 21935  
QY 233 CTTTACATTAATTTGACAGAAAGTGGATTGTGACGCGGACATTGAAATACAGTTTGC 352  
Db 21934 CTTTACATTAATTTGACAGAAAGTGGATTGTGACGCGGACATTGAAATACAGTTTGC 21875  
QY 353 AGGGGGATCGAAAGACACCAATCAGATGAAAGCCAGGAAGAGAGATGTACAGTT 412  
Db 21874 AGGGGGATCGAAAGACACCAATCAGATGAAAGCCAGGAAGAGAGATGTACAGTT 21815  
QY 413 CTTACGCTATGATATTTATGACAGATACAGCGTTCTTAAAGCCGAAAGTTTGAAGA 472  
Db 21814 CTTACGCTATGATATTTATGACAGATACAGCGTTCTTAAAGCCGAAAGTTTGAAGA 21755  
QY 473 GGAGATCCAGAGTGGTCTTTGATTACAACTATGAAGATCGTATAGTCTCTAGAAACA 532  
Db 21754 GGAGATCCAGAGTGGTCTTTGATTACAACTATGAAGATCGTATAGTCTCTAGAA-- 21697  
QY 533 GTAGACCGCATGGAAGACCCAGCGCGTAG--AGAAAGCATTCGACAAATGATAGACCAACT 591  
Db 21696 -TAGACCGCATGGAAGACCCAGCGCGTAG--AGAAAGCATTCGACAAATGATAGACCAACT 21638  
QY 592 GCAAGCTGGAATCCCACTACAGTTCGCTTACTACACTTCAGAAAGATCTGAAGCCGGA 651  
Db 21637 GCAAGCTGGAATCCCACTACAGTTCGCTTACTACACTTCAGAAAGATCTGAAGCCGGA 21578  
QY 652 AAAAGAACCAAGAAAGAGGAGTCAAGCGACCAAGAGGTGGTGAAGGTGCTGCAGTAT 711  
Db 21577 AAAAGAACCAAGAAAGAGGAGTCAAGCGACCAAGAGGTGGTGAAGGTGCTGCAGTAT 21518  
QY 712 GAATACCTGACGAATATTTTGAATTTGACTGCTGTAAGAAAGATTAAGATGTTATCGAAACT 771  
Db 21517 GAATACCTGACGAATATTTTGAATTTGACTGCTGTAAGAAAGATTAAGATGTTATCGAAACT 21458  
QY 772 ACATGGAATATTTGAAGTCCCTTCAGTTTGAAGTGAAGTATTTGGAACAATAAAG 831  
Db 21457 ACATGGAATATTTGAAGTCCCTTCAGTTTGAAGTGAAGTATTTGGAACAATAAAG 21398  
QY 832 AAATTAACCTTTGACTGTGGAACCTAATCCTTAATATGAATAGGTTTATATGATTC 891  
Db 21397 AAATTAACCTTTGACTGTGGAACCTAATCCTTAATATGAATAGGTTTATATGATTC 21338  
QY 892 ATGGGTAACAGGTCACTAATTAATTTATTTGAAACTAGAGTGTGAAATCAAGGAAGAC 951  
Db 21337 ATGGGTAACAGGTCACTAATTAATTTATTTGAAACTAGAGTGTGAAATCAAGGAAGAC 21278

QY 952 AGCATAGTCTCTTACAGTGCCTCTGTGGTCTGTCTCAAACTGAATTTGGTGGGAAAG 1011  
Db 21277 AGCATAGTCTCTTACAGTGCCTCTGTGGTCTGTCTCAAACTGAATTTGGTGGGAAAG 21218  
QY 1012 GTATGGTCCAAATTAAGAGTTCCATTTTGGCCATTTATGGCAATCTTGGCTTGTAT 1071  
Db 21217 GTATGGTCCAAATTAAGAGTTCCATTTTGGCCATTTATGGCAATCTTGGCTTGTAT 21158  
QY 1072 TTTGGTCCAGTGTCTTCTGCTTAATCATTTGCTTTGGCATCTGTTATTTACT 1131  
Db 21157 TTTGGTCCAGTGTCTTCTGCTTAATCATTTGCTTTGGCATCTGTTATTTACT 21098  
QY 1132 GTACACCAATGACAGTTTACATCTGCTTAACATCTCCCTCCAGGTAATTTCAATTA 1191  
Db 21097 GTACACCAATGACAGTTTACATCTGCTTAACATCTCCCTCCAGGTAATTTCAATTA 21038  
QY 1192 ATTTGACATCCGTAAGAGGCGCCATCTCTTCCACTCTTCTCCATGCTGATATCA 1251  
Db 21037 ATTTGACATCCGTAAGAGGCGCCATCTCTTCCACTCTTCTCCATGCTGATATCA 20978  
QY 1252 GCAAATATTTATTTAGCCCTTACTGTGGCAATCATTTGATGATTAATGAGAAAT 1311  
Db 20977 GCAAATATTTATTTAGCCCTTACTGTGGCAATCATTTGATGATTAATGAGAAAT 20918  
QY 1312 AGATTAATTCCTTATTTACATTAATGCTCTACTGACACATTTAGTGAATCTTACAGTAT 1371  
Db 20917 AGATTAATTCCTTATTTACATTAATGCTCTACTGACACATTTAGTGAATCTTACAGTAT 20858  
QY 1372 GGCTCATTTGTTTGTGAGTGTGTTATTCATTAACAATTTTACCCGTTGATCA 1431  
Db 20857 GGCTCATTTGTTTGTGAGTGTGTTATTCATTAACAATTTTACCCGTTGATCA 20798  
QY 1432 ATGTAATTTATGAACACATATTAAGATCAAGATTAATTTGTTGTTATTCGCAAT 1491  
Db 20797 ATGTAATTTATGAACACATATTAAGATCAAGATTAATTTGTTGTTATTCGCAAT 20738  
QY 1492 TAAAGATTCAGATTTTATGATCAATTAATTAATTAAGAAATGATTAATCTGTA 1551  
Db 20737 TAAAGATTCAGATTTTATGATCAATTAATTAATTAAGAAATGATTAATCTGTA 20678  
QY 1552 ATAACTGGTTTATTTGTCAGTGAATTAATTAATTAAGTAAATTTGTTACTC 1611  
Db 20677 ATAACTGGTTTATTTGTCAGTGAATTAATTAATTAAGTAAATTTGTTACTC 20618  
QY 1612 TGCCCTCAACAAACATGCTAGATATTAACCCCAAAATTAATTAATTTGATTAAG 1671  
Db 20617 TGCCCTCAACAAACATGCTAGATATTAACCCCAAAATTAATTAATTTGATTAAG 20558  
QY 1672 GTATTAAGAGACCTGGTGTCTAATTAATTAATTTTGAAGGACAGACAGAGCTGTATC 1731  
Db 20557 GTATTAAGAGACCTGGTGTCTAATTAATTAATTTTGAAGGACAGAGAGCTGTATC 20498  
QY 1732 CTAACTGATTTATGTTTCTGTAATTAAGAAATGTTACCAAAATTAATTAATTTAGTG 1791  
Db 20497 CTAACTGATTTATGTTTCTGTAATTAAGAAATGTTACCAAAATTAATTAATTTAGTG 20438  
QY 1792 ATTACATGATATTTATTAAGGACATGTTCTGTGTATTAAGCAATTAATTAATTTATTA 1851  
Db 20437 ATTACATGATATTTATTAAGGACATGTTCTGTGTATTAAGCAATTAATTAATTTATTA 20378  
QY 1852 GTA 1854  
Db 20377 GTA 20375

RESULT 8  
AY048592 16745 bp DNA linear PRI 26-SEP-2001  
LOCUS AY048592  
DEFINITION Homo sapiens TUS-associated SR proteins (TUSP) gene, complete cds,  
alternatively spliced.  
ACCESSION AY048592  
VERSION AY048592.1 GI:15787481  
KEYWORDS



[illegible]





QY 1362 ATTACAGTATGGCTCTTTTGTGTAGGCTGTATTCTATACAAATATTACACC 1421  
DB 36771 ATTACAGTATGGCTCTTTTGTGTAGGCTGTATTCTATACAAATATTACACC 36712  
QY 1422 ATTCGATCAATGATATTATAGACACATATACGATCAAGGATAGTATTGTGGTT 1481  
DB 36711 ATTCGATCAATGATATTATAGACACATATACGATCAAGGATAGTATTGTGGTT 36652  
QY 1482 ATTCGATCAATGATATTATAGACACATATATATTAATTAATGAATAATGAT 1541  
DB 36651 ATTCGATCAATGATATTATAGACACATATATTAATTAATGAATAATGAT 36592  
QY 1542 TTAATCTGTAATTAATCTGGTTATTGTGACGACGCTGTAATTAATTAATTAAT 1601  
DB 36591 TTAATCTGTAATTAATCTGGTTATTGTGACGACGCTGTAATTAATTAATTAAT 36532  
QY 1602 TTTGTTTACTGCTGCCCAACCAACACATGCTAGATATACCCCAAAATAGTATTAC 1661  
DB 36531 TTTGTTTACTGCTGCCCAACCAACACATGCTAGATATACCCCAAAATAGTATTAC 36472  
QY 1662 TTTGCTATTAGGATTAATAGAGAGAGCTGGGCTATTAATTAATTAATTAATTAAT 1721  
DB 36471 TTTGCTATTAGGATTAATAGAGAGAGCTGGGCTATTAATTAATTAATTAATTAAT 36412  
QY 1722 AGCTGTATCTCTTACTGATTTAGTATGTTCTGTATTAATTAATTAATTAATTAAT 1781  
DB 36411 AGCTGTATCTCTTACTGATTTAGTATGTTCTGTATTAATTAATTAATTAATTAAT 36352  
QY 1782 CTTTATTAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1841  
DB 36351 CTTTATTAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 36292  
QY 1842 AACTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1857  
DB 36291 AACTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 36276

RESULT 10  
AL590609/c 172307 bp DNA linear PRI 08-DEC-2003  
LOCUS Human DNA sequence from clone Rp11-4M23 on chromosome 1, complete  
DEFINITION  
AL590609  
ACCESSION  
AL590609.15 GI:39573520  
KEYWORDS  
HTG.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1 (bases 1 to 172307)  
Waller, J.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (08-DEC-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk; Clone requests: clonequests@sanger.ac.uk  
JOURNAL  
On Dec 8, 2003 this sequence version replaced gi:15795484.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: humquerry@sanger.ac.uk  
-----

as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emi, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMEP; Information on the WORMEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormep](http://www.sanger.ac.uk/Projects/C_elegans/wormep) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> Rp11-4M23 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBAC3.6.  
FEATURES  
source  
Location/Qualifiers  
1..172307  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="Rp11-4M23"  
/clone\_1id="RPCI-11.1"

ORIGIN  
Query Match 68.2%; Score 1276; DB 9; Length 172307;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 582 AGACCAACCTGCAGCTGGAATACCCAGTACAGTCTGCTTACTACCTCAAGAAAGATC 641  
DB 100747 AGACCAACCTGCAGCTGGAATACCCAGTACAGTCTGCTTACTACCTCAAGAAAGATC 100688  
QY 642 TGAAGGCGAAAAAGAACCAAGAGGAGGAGTTCAAGCAACCAAGGAGGAGTGAAGGT 701  
DB 100687 TGAAGGCGAAAAAGAACCAAGAGGAGGAGTTCAAGCAACCAAGGAGTGAAGGT 100628  
QY 702 GCTGCAGTATGAATCTGTAACGAATTTTGTACTCTGCTGTAAGAAAGATGTT 761  
DB 100627 GCTGCAGTATGAATCTGTAACGAATTTTGTACTCTGCTGTAAGAAAGATGTT 100568  
QY 762 ATCGAAATCTACATGGAATATGAAGTCCCTCAAGTTGAAGTGAAGATTTAGAC 821  
DB 100567 ATCGAAATCTACATGGAATATGAAGTCCCTCAAGTTGAAGTGAAGATTTAGAC 100508  
QY 822 AAATTAAGGAAATTCAACTTTGTACTTGTGAAACTAATCCCTAATATGAATAGTT 881  
DB 100507 AAATTAAGGAAATTCAACTTTGTACTTGTGAAACTAATCCCTAATATGAATAGTT 100448  
QY 882 ATATTGATTATGAGGTAACGGTTCATTAATTAATTTGGAACCTAGATGCTGAATAT 941  
DB 100447 ATATTGATTATGAGGTAACGGTTCATTAATTAATTTGGAACCTAGATGCTGAATAT 100388  
QY 942 CAAGAAAGACAGCATAGTCTCTTACAGTCCCTGTGTGCTGCTCAAGTGAATTTGG 1001  
DB 100387 CAAGAAAGACAGCATAGTCTCTTACAGTCCCTGTGTGCTGCTCAAGTGAATTTGG 100328  
QY 1002 GTGGGAAAAAGTATGCTCAATTAATTAATTAATTTGCAATTTGCAATATCTTGC 1061  
DB 100327 GTGGGAAAAAGTATGCTCAATTAATTAATTAATTTGCAATTTGCAATATCTTGC 100268  
QY 1062 CTTTGTATTATTTTGTGCGCAGTGTTCCTTAATCAATTTGCTTTGTGGCATCTGTGT 1121  
DB 100267 CTTTGTATTATTTTGTGCGCAGTGTTCCTTAATCAATTTGCTTTGTGGCATCTGTGT 100208  
QY 1122 TTATTACTTGTAACCAATGAGTGAATGATGCTTAATCAATCTTCCAGGATAA 1181  
DB 100207 TTATTACTTGTAACCAATGAGTGAATGATGCTTAATCAATCTTCCAGGATAA 100148  
QY 1182 TTCAATTATATTGACATCCAGTAAAGGAGGAGGAGTCTTCTGACCTTTCTTCAATGTC 1241

```

Db      100147 TTCCAAATTATTTGACATCCAGCTAAGAGGCCCATCTCTTCACCTCTTCTAGTC 100088
QY      1242 AGTATATTCAGCAAAATATTTATTGAGCCCTTACTGTGGGCAATCTGTAAGTAAT 1301
Db      100087 AGTATATTCAGCAAAATATTTATTGAGCCCTTACTGTGGGCAATCTGTAAGTAAT 100028
QY      1302 TGAGAAAAATAGATATTTCCCTTATTCAGTAATAGTCTACGAGACAACTAGTAGATC 1361
Db      100027 TGAGAAAAATAGATATTTCCCTTATTCAGTAATAGTCTACGAGACAACTAGTAGATC 99968
QY      1362 ATTACAGATGAGCTCATTTGTTTGTGAGAGTGTGTTATTCATACAAATATTTTACAC 1421
Db      99967 ATTACAGATGAGCTCATTTGTTTGTGAGAGTGTGTTATTCATACAAATATTTTACAC 99908
QY      1422 ATTGATCAATGATATTTATGAAACAATATAGCATAGAGATAGATATTTGTGTGTT 1481
Db      99907 ATTGATCAATGATATTTATGAAACAATATAGCATAGAGATAGATATTTGTGTGTT 99848
QY      1482 ATCTGCCATTTAAAGATATCCAGATTTGATCAGCTTATTTAAATTAAGAAAAATGAT 1541
Db      99847 ATCTGCCATTTAAAGATATCCAGATTTGATCAGCTTATTTAAATTAAGAAAAATGAT 99788
QY      1542 TTAATCTGTAATTAACCTGTTATTTGTCAGTACTGTAATATCTAGAGTTATATATA 1601
Db      99787 TTAATCTGTAATTAACCTGTTATTTGTCAGTACTGTAATATCTAGAGTTATATATA 99728
QY      1602 TTGTTTACTCTGCTCACCACCAACATGCTAGGATATTAACCCCAAAATAGTATTTAAC 1661
Db      99727 TTGTTTACTCTGCTCACCACCAACATGCTAGGATATTAACCCCAAAATAGTATTTAAC 99668
QY      1662 TTTCGATTAGGATTAAGAGAGCTGGTGTCTAATATTAATTTAGAGGAGACAGAG 1721
Db      99667 TTTCGATTAGGATTAAGAGAGCTGGTGTCTAATATTAATTTAGAGGAGACAGAG 99608
QY      1722 AGCTGTATCTTACTGATTTAGTATGTTCTGTAAATTTGAGAAATGTTTACCAAAATATA 1781
Db      99607 AGCTGTATCTTACTGATTTAGTATGTTCTGTAAATTTGAGAAATGTTTACCAAAATATA 99548
QY      1782 CTTTATGATTTACATGATATTTTATAGGGGACATGTTCTGTATAGGAATTAAT 1841
Db      99547 CTTTATGATTTACATGATATTTTATAGGGGACATGTTCTGTATAGGAATTAAT 99488
QY      1842 AACCTTTATAGTATCA 1857
Db      99487 AACCTTTATAGTATCA 99472

```

```

RESULT 11
HSM801468      2073 bp      mRNA      linear      PRI 18-FEB-2000
LOCUS      HSM801468
DEFINITION      Homo sapiens mRNA, cDNA DKFZp434N1717 (from clone DKFZp434N1717).
ACCESSION      AL133655
VERSION      AL133655.1 GI:6599245
KEYWORDS
SOURCE
ORGANISM      Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2073)
Blaecker,H., Boecher,M., Brandt,P., Mewes,H.W., Gaassenhuber,J. and
Wiemann,S.
Direct Submission
Submitted (15-DEC-1999) MIPS, Am Klopferpitz 18a, D-82152
Martinstedt, GERMANY
COMMENT
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by GPF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp434N1717) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available

```

```

FEATURES
source
at http://www.mips.biochem.mpg.de/proj/cDNA/.
location/Qualifiers
1..2073
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="716_F_2; 850_F_7; 882_B_6; 900_C_1; 912_B_9;
964_H_2; 800_D_12; 809_H_11; (934,936)_C_10"
/clone="DKFZp434N1717"
/tissue_type="testis"
/clone_id="434 (synonym: htes3). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
2001..2006
polya_signal
polya_site
2016
ORIGIN

```

Query Match 68.0%; Score 1273; DB 9; Length 2073;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      582 AGACCAACTGCACTGCAATGATACCAAGTCTGCTTACTACACTTCAAGAAAGATC 641
Db      748 AGACCAACTGCACTGCAATGATACCAAGTCTGCTTACTACACTTCAAGAAAGATC 807
QY      642 TGAAAGCGGAAAAAGAACCAAGGAGGAGTCAAGGACCAAGGAGGAGTGAAGGT 701
Db      808 TGAAAGCGGAAAAAGAACCAAGGAGGAGTCAAGGACCAAGGAGGAGTGAAGGT 867
QY      702 GCTGCAATGATATCTGATGATATTTTGAATCTGCTGTAAGATTAAGATGAT 761
Db      868 GCTGCAATGATATCTGATGATATTTTGAATCTGCTGTAAGATTAAGATGAT 927
QY      762 ATCCGAAACTCATGGAATTAATTTGAAGTCCCTCAAGTTGAAAGTAAGATTTAGAGC 821
Db      928 ATCCGAAACTCATGGAATTAATTTGAAGTCCCTCAAGTTGAAAGTAAGATTTAGAGC 987
QY      822 AAATTAAGGAAATTTCACTTTGTAATTTGGAATTTGGAATTTGGAATTTGGAATTT 881
Db      988 AAATTAAGGAAATTTCACTTTGTAATTTGGAATTTGGAATTTGGAATTTGGAATTT 1047
QY      882 ATATTGATTCATGGGTAAACAGGTCTATTAATTTGGAATTTGGAATTTGGAATTT 941
Db      1048 ATATTGATTCATGGGTAAACAGGTCTATTAATTTGGAATTTGGAATTTGGAATTT 1107
QY      942 CAAGGAAGACAGCCATGATGCTCTTACAGGCTCTGTTGGTCTGCTCAAGTGAATTC 1001
Db      1108 CAAGGAAGACAGCCATGATGCTCTTACAGGCTCTGTTGGTCTGCTCAAGTGAATTC 1167
QY      1002 GTGGGAAAAAGTATGATCCATATTAAGATTTTCCATTTTGGCATTTTGGCAATCTTGC 1061
Db      1168 GTGGGAAAAAGTATGATGCTCCATATTAAGATTTTCCATTTTGGCATTTTGGCAATCTTGC 1227
QY      1062 CTTTGTATTTTGTGTGCGAGGTTTTCGCTTAATCAATTTGCTTTGGGATCTGTGT 1121
Db      1228 CTTTGTATTTTGTGTGCGAGGTTTTCGCTTAATCAATTTGCTTTGGGATCTGTGT 1287
QY      1122 TTATTACTTGTACACACATGCAAGTTTACATCTGCTTAAGTACTGCTTCCAGGTAAA 1181
Db      1288 TTATTACTTGTACACACATGCAAGTTTACATCTGCTTAAGTACTGCTTCCAGGTAAA 1347
QY      1182 TTCCAAATTATTTGACATCCAGCTAAGAGGCCCATCTCTTCTCACTTTCTTAGTC 1241
Db      1348 TTCCAAATTATTTGACATCCAGCTAAGAGGCCCATCTCTTCTCACTTTCTTAGTC 1407
QY      1242 AGTATATTCAGCAAAATATTTATTGAGCCCTTACTGTGGGCAATCTGTAAGTAAT 1301
Db      1408 AGTATATTCAGCAAAATATTTATTGAGCCCTTACTGTGGGCAATCTGTAAGTAAT 1467
QY      1302 TGAGAAAAATAGATATTTCCCTTATTCAGTAATAGTCTACTGAGCACAATCTAGTAGATC 1361
Db      1468 TGAGAAAAATAGATATTTCCCTTATTCAGTAATAGTCTACTGAGCACAATCTAGTAGATC 1527

```

QY 1362 ATTACAGTATGGCCCTCATTTGTTTGAAGTGTGTTATTCATTAACATATTTTACACC 1421  
DB 1528 ATTACAGTATGGCCCTCATTTGTTTGAAGTGTGTTATTCATTAACATATTTTACACC 1587  
QY 1422 ATTCGTATCAATGTATTTATTAAGACACATATATACATCAAGGATAGTATTTGTTGGTT 1481  
DB 1588 ATTCGTATCAATGTATTTATTAAGACACATATATACATCAAGGATAGTATTTGTTGGTT 1647  
QY 1482 ATTCGCCATTTAAAGTATCCAGTATTTGATACATTTATTAATTAATGAATAAATGAT 1541  
DB 1648 ATTCGCCATTTAAAGTATCCAGTATTTGATACATTTATTAATTAATGAATAAATGAT 1707  
QY 1542 TTAATCTGTAAATACTGGTTTATTTGTGACAGTACCTGTAAATCTAGATTTATTAATA 1601  
DB 1708 TTAATCTGTAAATACTGGTTTATTTGTGACAGTACCTGTAAATCTAGATTTATTAATA 1767  
QY 1602 TTGTTTACTCTGCCCTCACCAACACATGCTAGATATTAACCCCAAAATTAAGTATTATAC 1661  
DB 1768 TTGTTTACTCTGCCCTCACCAACACATGCTAGATATTAACCCCAAAATTAAGTATTATAC 1827  
QY 1662 TTTTGATTAAGTATTAAGGAGACTGGGTCTATTAATTAATTTTGAAGGACAGACAG 1721  
DB 1828 TTTTGATTAAGTATTAAGGAGACTGGGTCTATTAATTAATTTTGAAGGACAGACAG 1887  
QY 1722 AGCTGTATCCCTAACTGATTTAGTATGTCTGTAAATTAAGAAATAGTTTCAACAAATTATA 1781  
DB 1888 AGCTGTATCCCTAACTGATTTAGTATGTCTGTAAATTAAGAAATAGTTTCAACAAATTATA 1947  
QY 1782 CTTTATTAAGTATTAAGTATTAAGGAGCATGTTCTGTGTATAGCAATTAAT 1841  
DB 1948 CTTTATTAAGTATTAAGTATTAAGGAGCATGTTCTGTGTATAGCAATTAAT 2007  
QY 1842 AACTTTTATAGTA 1854  
DB 2008 AACTTTTATAGTA 2020

RESULT 12  
AK093226  
LOCUS 2834 bp mRNA linear PRI 15-JUL-2002  
DEFINITION Homo sapiens cDNA FLJ35907 fls, clone TESR12009785, weakly similar to Mus musculus TLS-associated protein with SR repeats mRNA.  
ACCESSION AK093226  
VERSION AK093226.1 GI:21752034  
KEYWORDS oligo capping, fls (full insert sequence).  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS 1 Iihabashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S., Ono, Y., Hotuta, T., Hiraoka, S., Murakawa, K., Takiguchi, S., Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M., Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Magatsu, M., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagatairi, K., Masuno, Y., Nagai, K. and Isogai, T.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2834)  
AUTHORS Isogai, T. and Yamamoto, J.  
TITLE Direct Submision  
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazuoka-Kametari, Kisarazu, Chiba 282-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and

FEATURES  
source  
location/Qualifiers  
1..2834  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="TESR12009785"  
/tissue\_type="testis"  
/clone\_lib="TESR12"  
/note="Cloning vector: pME185FLJ3"

ORIGIN  
Query Match 65.4%; Score 1225; DB 9; Length 2834;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1565; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

Evaluation: clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

QY 290 ACGCTTTACATTAATTTGACAGAAAGTGGATTTTGACAGGCAATGGAATGAGTTTG 349  
DB 1266 ACGCTTTACATTAATTTGACAGAAAGTGGATTTTGACAGGCAATGGAATGAGTTTG 1325  
QY 350 CCCAGGGGGAATCGAAAGACACCAATCATGATGAAAGCCAGAGAGAGAAATGTGTACA 409  
DB 1326 CCCAGGGGGAATCGAAAGACACCAATCATGATGAAAGCCAGAGAGAGAAATGTGTACA 1385  
QY 410 GTTCTTCACGCTATGATGATTTATGACAGATACAGACGTTCTTAAGAACCGGAATTTGAAA 469  
DB 1386 GTTCTTCACGCTATGATGATTTATGACAGATACAGACGTTCTTAAGAACCGGAATTTGAAA 1445  
QY 470 GGAGAGATCAAGAGTGTGCTTTTGAATTAACAATAGAAAGATCGTATAGTCTAGAA 529  
DB 1446 GGAGAGATCAAGAGTGTGCTTTTGAATTAACAATAGAAAGATCGTATAGTCTAGAA 1505  
QY 530 ACAATTAACCGACATCGAAAGACACCGGCTAG-AGAAGCATTCGACATGATAGACCA 588  
DB 1506 ACAATTAACCGACATCGAAAGACACCGGCTAGACCAATTCGACATGATAGACCA 1565  
QY 589 ACTGACGCTGGAATATACCAAGTACGCTGCTTAATCTCACTTCAAGAAAGATCTGAAGC 648  
DB 1566 ACTGACGCTGGAATATACCAAGTACGCTGCTTAATCTCACTTCAAGAAAGATCTGAAGC 1625  
QY 649 GGAAGAAAGAAACCAAGAGGAGGAGTTCAGACCAAGAGGTGGAGAGGTCTGCAG 708  
DB 1626 GGAAGAAAGAAACCAAGAGGAGGAGTTCAGACCAAGAGGTGGAGAGGTCTGCAG 1685  
QY 709 TATGAATCTGTACGAATATTTTGACTCTGCTGAGAAAGATTAAGAAATGTTATCGAAA 768  
DB 1686 TATGAATCTGTACGAATATTTTGACTCTGCTGAGAAAGATTAAGAAATGTTATCGAAA 1745  
QY 769 ACTACATGGAATTAATGAAGTCCCTTCAAGTTTGAAGTAAAGTATTTAGACCAATTA 828  
DB 1746 ACTACATGGAATTAATGAAGTCCCTTCAAGTTTGAAGTAAAGTATTTAGACCAATTA 1805  
QY 829 AGGAAATTCACCTTTGACTTTGGAACCTAATCCCTAATATGAATAGAGTTATATTTGA 888  
DB 1806 AGGAAATTCACCTTTGACTTTGGAACCTAATCCCTAATATGAATAGAGTTATATTTGA 1865  
QY 889 TTATATGGGTACAGGTTCATTAATTAATTTATGGAACCTAGAGTGTGAATATCAAGAA 948  
DB 1866 TTATATGGGTACAGGTTCATTAATTAATTTATGGAACCTAGAGTGTGAATATCAAGAA 1924  
QY 949 GACAGCATAGTGTCTTACAGTGCCTGTGTGTCTGTCTCAACATGAATTTGGTGGAA 1008  
DB 1925 GACAGCATAGTGTCTTACAGTGCCTGTGTGTCTGTCTCAACATGAATTTGGTGGAA 1984  
QY 1009 AAGGATATGATCAATTAATTAAGTTCATTTTGGCAATATTTGCAATCTTGCTTTGTT 1068  
DB 1985 AAGGATATGATCAATTAATTAAGTTCATTTTGGCAATATTTGCAATCTTGCTTTGTT 2044  
QY 1069 TATTTTGGTCCAGTGTCTTGTCTTAATCATATTTGCTTTGTTGGCATCTGTGTTATTTA 1128  
DB 2045 TATTTTGGTCCAGTGTCTTGTCTTAATCATATTTGCTTTGTTGGCATCTGTGTTATTTA 2104

QY 1129 CTGTACACACATGACGATTACATCTGTCTTAACTCTCTCCAGGTAATTCAT 1188  
 1129 |||||  
 Db 2105 CTGTACACACATGACGATTACATCTGTCTTAACTCTCTCCAGGTAATTCAT 2164  
 2105 |||||  
 QY 1189 TATATTGACATCCAGTAAGAGGCCCCATCTCTTCACCTCTTCTCTAGTATAT 1248  
 1189 |||||  
 Db 2165 TATATTGACATCCAGTAAGAGGCCCCATCTCTTCACCTCTTCTCTAGTATAT 2224  
 2165 |||||  
 QY 1249 TCAGCAAAATATTTATGAGCCCTTACTGTGGCAATCATTTGATTAATTAAGAAA 1308  
 1249 |||||  
 Db 2225 TCAGCAAAATATTTATGAGCCCTTACTGTGGCAATCATTTGATTAATTAAGAAA 2284  
 2225 |||||  
 QY 1309 AATGATTAATTCCTTATTCGTAATATGTCTACGACACATCTAGTGATTCATTACG 1368  
 1309 |||||  
 Db 2285 AATGATTAATTCCTTATTCGTAATATGTCTACGACACATCTAGTGATTCATTACG 2344  
 2285 |||||  
 QY 1369 TATGGCCCATTTGTTGTTGAGGTGTTTATCATACAAATTTTACACCATTCGTA 1428  
 1369 |||||  
 Db 2345 TATGGCCCATTTGTTGTTGAGGTGTTTATCATACAAATTTTACACCATTCGTA 2404  
 2345 |||||  
 QY 1429 TCAATGTAATTAAGAACACATATACGATCAAGGATTAATTTGTGTATCTGCC 1488  
 1429 |||||  
 Db 2405 TCAATGTAATTAAGAACACATATACGATCAAGGATTAATTTGTGTATCTGCC 2464  
 2405 |||||  
 QY 1489 ATTAAAAAGTATCCAGTATTTGATCACTTATTAATTAATGAATAATGATTTAATCT 1548  
 1489 |||||  
 Db 2465 ATTAAAAAGTATCCAGTATTTGATCACTTATTAATTAATGAATAATGATTTAATCT 2524  
 2465 |||||  
 QY 1549 GTAATTAACCTGGTTATTTGTCAGTACGTAATATACAGTATTAATTAATTTGTTA 1608  
 1549 |||||  
 Db 2525 GTAATTAACCTGGTTATTTGTCAGTACGTAATATACAGTATTAATTAATTTGTTA 2584  
 2525 |||||  
 QY 1609 CTCTGCTCCACCAACACATGCTAGATATAACCCCCAAATAGTATTTAACTTGCAT 1668  
 1609 |||||  
 Db 2585 CTCTGCTCCACCAACACATGCTAGATATAACCCCCAAATAGTATTTAACTTGCAT 2644  
 2585 |||||  
 QY 1669 TAGGTATTAAGAGACCTGGGTCTCTTAATTAATTTTGGGACAGAGACCTGTT 1728  
 1669 |||||  
 Db 2645 TAGGTATTAAGAGACCTGGGTCTCTTAATTAATTTTGGGACAGAGACCTGTT 2704  
 2645 |||||  
 QY 1729 ATCCTAACGATTTGATGTTCTGTAAATTGAGAAATGTCACCAATTTACTTTTAA 1788  
 1729 |||||  
 Db 2705 ATCCTAACGATTTGATGTTCTGTAAATTGAGAAATGTCACCAATTTACTTTTAA 2764  
 2705 |||||  
 QY 1789 GTGATTAATGATTAATTTATAGGAGACATGTTCTGTATAGCGAATTAATCTTTT 1848  
 1789 |||||  
 Db 2765 GTGATTAATGATTAATTTATAGGAGACATGTTCTGTATAGCGAATTAATCTTTT 2824  
 2765 |||||  
 QY 1849 AATGATTA 1857  
 1849 |||||  
 Db 2825 AATGATTA 2833  
 2825 |||||  
 RESULT 13  
 BD23773 1326 bp DNA linear PAT 17-JUL-2003  
 LOCUS Human RNA-associated proteins.  
 DEFINITION BD23773.1 GI:3303543  
 ACCESSION BD23773.1 GI:3303543  
 VERSION JP 2002523045-A/9  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiindae; Homo.  
 1 (bases 1 to 1326)  
 Hillman, J.L., Yue, H., Tang, Y.T., Corley, N.C., Guegler, K.J.,  
 Gorgone, G.A., Patterson, C., Baughn, M.R., Lal, P., Bandman, O.,  
 Reddy, R., Azimzai, Y., Shih, L.L., Yang, J., and Lu, D.A.M.  
 Human RNA-associated proteins  
 Patent: JP 2002523045-A 9 30-JUL-2002;  
 INCYTE PHARMACEUTICALS INC  
 OS Homo sapiens (human)  
 COMMENT PN JP 2002523045-A/9

PD 30-JUL-2002  
 PF 20-AUG-1999 JP 2000566425  
 ER 21-AUG-1998 US 60/097550, 12-JAN-1999 US 60/115639 PI  
 JENNIFER L HILLMAN, HENRY YUE, Y TOM TANG, NEIL C CORLEY, KARL J PI  
 GUEGLER,  
 PI GINA A GORGONE, CHANDRA PATTERSON, MARIAN R BAUGHN, PREETI TAL,  
 PI OLGA BANDMAN,  
 PI ROOPA REDDY, YALDA AZIMZAI, LEO L SHIH, JUNGMIN  
 YANG, DYONG AINA M  
 YI LU  
 PC C12N15/09, A61K38/00, A61K38/55, A61K45/00, A61P1/00, A61P1/04, PC  
 A61P1/16,  
 PC A61P1/18, A61P3/00, A61P3/10, A61P7/00, A61P9/00, A61P11/  
 PC 00, A61P11/06,  
 PC A61P13/12, A61P17/00, A61P17/06, A61P19/00, A61P19/02, A61P19/06,  
 PC A61P19/10,  
 PC A61P21/00, A61P21/04, A61P29/00, A61P29/00, A61P31/00, A61P31/04,  
 PC A61P31/10,  
 PC A61P31/12, A61P31/18, A61P31/22, A61P35/00, A61P35/02, A61P37/00,  
 PC A61P37/08,  
 PC C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10 PC  
 , C12P21/02, C12Q1/68,  
 PC G01N33/15, G01N33/50, G01N33/53, G01N33/566//C12P21/08, C12N15/00,  
 PC C12N5/00,  
 PC A61K37/02, A61K37/64  
 PC A61K37/02, A61K37/64  
 CC a or t or g or c, unknown, or other  
 CC Incyte Identification No.: 934406CB1  
 FH Key Location/Qualifiers  
 FT unsure 1313.  
 Location/Qualifiers  
 1..1326  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

ORIGIN  
 Query Match 50.2%; Score 940; DB 6; Length 1326;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1180; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 GCCCTTAGCTGTGTGTCTGAGGCTGGCCGCTGAGCCGCGGACGGTTTGTGAGGC 60  
 Db 129 GCCCTTAGCTGTGTGTCTGAGGCTGGCCGCTGAGCCGCGGACGGTTTGTGAGGC 188  
 129 |||||  
 QY 61 CGTTAGTGGCCCGCCGAGACAGCCCGCGCATGCGCTACCTGGTCCCGCCCAAC 120  
 61 |||||  
 Db 189 CGTTAGTGGCCCGCCGAGACAGCCCGCGCATGCGCTACCTGGTCCCGCCCAAC 248  
 189 |||||  
 QY 121 ACGTCTGTGTGTGTGAGAACGTGGCCGAGACAGACAGGTCTGAAGACTTGGCGGCTGAA 180  
 121 |||||  
 Db 249 ACGTCTGTGTGTGTGAGAACGTGGCCGAGACAGACAGGTCTGAAGACTTGGCGGCTGAA 308  
 249 |||||  
 QY 181 TTTGGTGTATGTGTCTATAGTTGATGTGTATGTTTCCACTTGATTTCTACGCTGCGCT 240  
 181 |||||  
 Db 309 TTTGGTGTATGTGTCTATAGTTGATGTGTATGTTTCCACTTGATTTCTACGCTGCGCT 368  
 309 |||||  
 QY 241 CCAAGAGATTTGCTATGTTCAATTTGAGATGTTGCGATGCTGAAGACGCTTTACAT 300  
 241 |||||  
 Db 369 CCAAGAGATTTGCTATGTTCAATTTGAGATGTTGCGATGCTGAAGACGCTTTACAT 428  
 369 |||||  
 QY 301 AATTGGACAGAAAGTGAATTTGTGGACGGCAGATTGAATACAGTTTGGCCAGGGGAT 360  
 301 |||||  
 Db 429 AATTGGACAGAAAGTGAATTTGTGGACGGCAGATTGAATACAGTTTGGCCAGGGGAT 488  
 429 |||||  
 QY 361 CGAAGACACCAATATAGATGAAGAGCCAGAGAAAGAGAGAAAGTGTACGTTCTTACGC 420  
 361 |||||  
 Db 489 CGAAGACACCAATATAGATGAAGAGCCAGAGAAAGAGAGAAAGTGTACGTTCTTACGC 548  
 489 |||||  
 QY 421 TATGATGATTAAGACAGATACAGACGTTCTGAAGCCGAAGTTATGAAAGAGAGATCA 480  
 421 |||||  
 Db 549 TATGATGATTAAGACAGATACAGACGTTCTGAAGCCGAAGTTATGAAAGAGAGATCA 608  
 549 |||||  
 QY 481 AGAAGTCGCTCTTTGATTAACAATATAGAAAGTCGTATAGCTCTAGAAAACGTTACCG 540  
 481 |||||

Db 609 AAAAAAGCGCTCTTTGATTACAACTATAGAGATGCTATATGCTTAGAAAACGTAGACCG 668  
Qy 541 ACTGAAAGACCAACGGCGTAG-AGAAAGCAATTCGCAAAATGATAGACCAAACTGACGTGG 599  
Db 669 ACTGAAAGACCAACGGCGTAGAGAAAGCAATTCGCAAAATGATAGACCAAACTGACGTGG 728  
Qy 600 AATACCCAGTACAGTTCTGCTTACTACACTTCAAGAAAGATCTGAAAGCGGAAAAAGAAC 659  
Db 729 AATACCCAGTACAGTTCTGCTTACTACACTTCAAGAAAGATCTGAAAGCGGAAAAAGAAC 788  
Qy 660 CAAAGAGGGCAGTTCAGAGCACAAGGGGTGGGAAAGGTCGTGCAATGATAGATCTG 719  
Db 769 CAAAGAGGGCAGTTCAGAGCACAAGGGGTGGGAAAGGTCGTGCAATGATAGATCTG 848  
Qy 720 TAGGAATAATTTGACTCTGCTGTAAGAAAGATGAAGATGTTATGCAAACTACATGGAA 779  
Db 849 TACGAATAATTTGACTCTGCTGTAAGAAAGATGAAGATGTTATGCAAACTACATGGAA 908  
Qy 780 TAATTGAAGTCCCTTCAAGTTTGAAGATGATGATTTAGCAAAATTAAGAAATTCGA 839  
Db 909 TAATTGAAGTCCCTTCAAGTTTGAAGATGATGATTTAGCAAAATTAAGAAATTCGA 968  
Qy 840 CTTTGACTGTGGAACCTAATCCCTAATATGAATAGTTTATATGATTCATGGGTAA 899  
Db 969 CTTTGACTGTGGAACCTAATCCCTAATATGAATAGTTTATATGATTCATGGGTAA 1028  
Qy 900 CAGGTCCATTAATAATTAATTTGAAACCTAGATGCTGAAATCAAGAAAGACAGCATAG 959  
Db 1029 CAGGTCCATTAATAATTAATTTGAAACCTAGATGCTGAAATCAAGAAAGACAGCATAG 1088  
Qy 960 TCTCTTACAGTCCCTTCTGTTGCTGCTCTCAAACTGAAATGGGTGGGAAAAAGTATGTC 1019  
Db 1089 TCTCTTACAGTCCCTTCTGTTGCTGCTCTCAAACTGAAATGGGTGGGAAAAAGTATGTC 1148  
Qy 1020 CAATATAAAGATGTCATTTTGGCATTAAT-GGCAATCTGCGTTGTTATTTGGTG 1078  
Db 1149 CAATATAAAGATGTCATTTTGGCATTAAT-GGCAATCTGCGTTGTTATTTGGTG 1208  
Qy 1079 CCAAGTCTTTCTGCTTAATCAATTTGCTTTGTCATCTGTTTATTTATCTGTACAC 1138  
Db 1209 CCAAGTCTTTCTGCTTAATCAATTTGCTTTGTCATCTGTTTATTTATTTATCTGTACAC 1268  
Qy 1139 ACATGAGTTTACATCTGCTTACTACTCTCTTCCAGGTAA 1180  
Db 1269 ACATGAGTTTACATCTGCTTACTACTCTCTTCCAGGTAA 1310

RESULT 14  
AX713349 2473 bp DNA linear PAT 15-APR-2003  
LOCUS AX713349  
DEFINITION Sequence 33 from Patent EP1293569.  
ACCESSION AX713349  
VERSION AX713349.1 GI:29888205  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Iosagi,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,  
Yamanoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,  
Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and  
Masuko,Y.  
TITLE Full-length cDNAs  
JOURNAL Patent: EP 1293569-A 33 19-MAR-2003;  
Helix Research Institute (UP) ; Research Association for  
Biotechnology (UP)  
FEATURES  
Source  
1. .2473  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 42.0%; Score 786; DB 6; Length 2473;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 836; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 582 AGACCAAACTGAGCTGGAATACCGATGACAGTCTGCTTACTACACTTCAAGAAAGTC 641  
Db 1388 AGACCAAACTGAGCTGGAATACCGATGACAGTCTGCTTACTACACTTCAAGAAAGTC 1447  
Qy 642 TGAAGCCGAAAAAGAACCAAGAAAGGCGAGTTCAAGGACCAAGAGGTTGGTGAAGGT 701  
Db 1448 TGAAGCCGAAAAAGAACCAAGAAAGGCGAGTTCAAGGACCAAGAGGTTGGTGAAGGT 1507  
Qy 702 GGTGAGATGAATATCTGTACGAATATTTTGAATCTGCTGTAAGAAAGTAAAGATGTT 761  
Db 1508 GGTGAGATGAATATCTGTACGAATATTTTGAATCTGCTGTAAGAAAGTAAAGATGTT 1567  
Qy 762 ATCGAAACCTACATGAAATATTAAGTCCCTCAAGTTTGAAGTAAGCAATTTAGAGC 821  
Db 1568 ATCGAAACCTACATGAAATATTAAGTCCCTCAAGTTTGAAGTAAGCAATTTAGAGC 1627  
Qy 822 AAATTAAGAAATTCGAATCTTTGTAATTTGGAACCTAATCCCTAATATGAAATAGTTC 881  
Db 1628 AAATTAAGAAATTCGAATCTTTGTAATTTGGAACCTAATCCCTAATATGAAATAGTTC 1687  
Qy 882 ATATTGATTCATGGGTAAACAGTCCATTAATAATTTTGGAACTAGGATGCTGAATAT 941  
Db 1688 ATATTGATTCATGGGTAAACAGTCCATTAATAATTTTGGAACTAGGATGCTGAATAT 1747  
Qy 942 CAAAGAAAGACGCAATAGTCTCTTAAGTCCCTGTTGCTGCTCTCAAACTGAATGG 1001  
Db 1748 CAAAGAAAGACGCAATAGTCTCTTAAGTCCCTGTTGCTGCTCTCAAACTGAATGG 1807  
Qy 1002 GTGGGAAAAAGTATGCTCAATATAAAGTTTCAATTTTGGCAATATTTGGCAAACTCTTG 1061  
Db 1808 GTGGGAAAAAGTATGCTCAATATAAAGTTTCAATTTTGGCAATATTTGGCAAACTCTTG 1867  
Qy 1062 CTTTGTTATTTTGGTGGCCAGTCTTCTGCTTAATCAATTTGCTTTGTTGGCATCTGCT 1121  
Db 1868 CTTTGTTATTTTGGTGGCCAGTCTTCTGCTTAATCAATTTGCTTTGTTGGCATCTGCT 1927  
Qy 1122 TTATTTACTGTACACCAATGCAATGCAATCTGCTTAATCAATCTCTTCCAGGTAA 1181  
Db 1928 TTATTTACTGTACACCAATGCAATGCAATCTGCTTAATCAATCTCTTCCAGGTAA 1987  
Qy 1182 TTCCAAATTAATTTGACATCCAGTAAAGGCGCCATCTCTTCCACCTCTTCCATGTC 1241  
Db 1988 TTCCAAATTAATTTGACATCCAGTAAAGGCGCCATCTCTTCCACCTCTTCCATGTC 2047  
Qy 1242 AGTAAATTTGCAAAATTTTATTAAGGCGCTTACTGTTGGGCAAAATCATTTGATGATTAAT 1301  
Db 2048 AGTAAATTTGCAAAATTTTATTAAGGCGCTTACTGTTGGGCAAAATCATTTGATGATTAAT 2107  
Qy 1302 TGAAGAAATTAATTAATTTCCCTTATTCAGTAATATGCTACTGAGCAATCTAGTAAATC 1361  
Db 2108 TGAAGAAATTAATTAATTTCCCTTATTCAGTAATATGCTACTGAGCAATCTAGTAAATC 2167  
Qy 1362 ATTACAGTATGCGCTCATCTGTTGTTGAGGTGCTGTTATTCATTAATATATTTTAC 1418  
Db 2168 ATTACAGTATGCGCTCATCTGTTGTTGAGGTGCTGTTATTCATTAATATATTTTAC 2224

RESULT 15  
AK054635 2473 bp mRNA linear PRI 01-AUG-2002  
LOCUS AK054635  
DEFINITION Homo sapiens cDNA FLJ30073 fis, clone ASTRO2000480.  
ACCESSION AK054635  
VERSION AK054635.1 GI:16549217  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



REFERENCE  
AUTHORS  
1 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
Oshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K.,  
Arita, M., Mushashi, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R.,  
Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,  
Kawai, Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,  
Masuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,  
Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B.,  
Suzuki, Y., Sugano, S., Nagahori, K., Masuo, Y., Nagai, K. and  
Isozaki, T.  
TITLE  
JOURNAL  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 2473)  
Isozaki, T., Otsuki, T. and Sugiyama, T.  
REFERENCE  
AUTHORS  
JOURNAL  
TITLE  
Direct Submission  
Submitted (24-OCT-2001) Takao Isozaki, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail: genomics@hri.co.jp; Tel: 81-438-52-3975; Fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: RAB and  
HRI.  
FEATURES  
source  
1..2473  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="ASTRO200480"  
/cell\_type="normal astrocytes (NHA5732)"  
/clone\_lib="ASTRO2"  
/note="Cloning vector: pME18FL3-primary culture, normal  
astrocytes"

ORIGIN  
Query Match 42.0%; Score 786; DB 9; Length 2473;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 836; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 562 AGACCAAACTGCAAGTGGAAATACCCAGTACAGTTCCTTACTACACTCAAGAAAGATC 641  
DB 1388 AGACCAAACTGCAAGTGGAAATACCCAGTACAGTTCCTTACTACACTCAAGAAAGATC 1447  
QY 642 TGAAGCGGAAAAAGAACCAAGAGGAGGAGTTCAAGCAGCAAGAGGAGTGGAGAGGT 701  
DB 1448 TGAAGCGGAAAAAGAACCAAGAGGAGGAGTTCAAGCAGCAAGAGGAGTGGAGAGGT 1507  
QY 702 GCTGCAGTATGAATACCTGTAGCAATATTTTGACTGTGCTGTGAAAAAGATAAAAAGATGTT 761  
DB 1508 GCTGCAGTATGAATACCTGTAGCAATATTTTGACTGTGCTGTGAAAAAGATAAAAAGATGTT 1567  
QY 762 ATCGAAAACTACATGAAATTAATTGAATGCCCTTCAAGTTGAAAGTAAGCATTTTAGAGAC 821  
DB 1568 ATCGAAAACTACATGAAATTAATTGAATGCCCTTCAAGTTGAAAGTAAGCATTTTAGAGAC 1627  
QY 822 AATATTAAGGAATTCACCTTGTACTTGTGAAACTAATCCCTAATATGAATAGTTT 881  
DB 1628 AATATTAAGGAATTCACCTTGTACTTGTGAAACTAATCCCTAATATGAATAGTTT 1687  
QY 882 AATATGATTCATGGGTACAGGTCCAAATTAATTAATTGAAAACTAGAGTGTCTGAATAT 941  
DB 1688 AATATGATTCATGGGTACAGGTCCAAATTAATTAATTGAAAACTAGAGTGTCTGAATAT 1747  
QY 942 CAAGGAGACAGCAGTACAGTCTCTTACAGTGCCTGTGTGCTGTCTCAACAGTAATGG 1001  
DB 1748 CAAGGAGACAGCAGTACAGTCTCTTACAGTGCCTGTGTGCTGTCTCAACAGTAATGG 1807  
QY 1002 GTGGGAAAAAGTATGGTCCAAATTAATTAAGTTCATTTTGCATTTATTTGGCAATCTTGC 1061  
DB 1808 GTGGGAAAAAGTATGGTCCAAATTAATTAAGTTCATTTTGCATTTATTTGGTAAATCTTGC 1867

QY 1062 CTTTGTATTTTGGTGGCAGGTGTTTCTGTATATCATTTTGTGTGGCATCTGNGT 1121  
DB 1868 CTTTGTATTTTGGTGGCAGGTGTTTCTGTATATCATTTTGTGTGGCATCTGNGT 1927  
QY 1122 TTATTTCTTTACACACAGCAGTATTTACATCTGTCTTACTACTCTCTTCCAGGTAAA 1181  
DB 1928 TTATTTCTTTACACACAGCAGTATTTACATCTGTCTTACTACTCTCTTCCAGGTAAA 1987  
QY 1182 TTCCAAATTAATTTGACATCCAGCTAAGAGGAGGCCATCTCTTCTACCTCTTTCTAGTC 1241  
DB 1988 TTCCAAATTAATTTGACATCCAGCTAAGAGGAGGCCATCTCTTCTACCTCTTTCTAGTC 2047  
QY 1242 AGTATATTCACCAATATTTTATTTAGGCCCTTACAGTGGGCAATTCATTTGACTGGATAT 1301  
DB 2048 AGTATATTCACCAATATTTTATTTAGGCCCTTACAGTGGGCAATTCATTTGACTGGATAT 2107  
QY 1302 TGAGAAAAATAGATTAATCCCTTATTCAGTAAATGTCTACTGAGCACAATCTAGTGAATC 1361  
DB 2108 TGAGAAAAATAGATTAATCCCTTATTCAGTAAATGTCTACTGAGCACAATCTAGTGAATC 2167  
QY 1362 ATTACAGTATGAGCTCATTTGTTTGTGAGGTGTATTCATACAAATTTTAC 1418  
DB 2168 ATTACAGTATGAGCTCATTTGTTTGTGAGGTGTATTCATACAAATTTTAC 2224

Search completed: July 29, 2004, 07:58:23  
Job time : 7426 secs





XX 23-MAR-2000.  
PD  
XX  
XX 17-SEP-1999; 99WO-US021688.  
PF  
XX 17-SEP-1998; 98US-00156039.  
PR 22-SEP-1998; 98US-00158720.  
PR 04-NOV-1998; 98US-00186815.  
PR 08-APR-1999; 99US-0128660P.  
XX  
XX (INCY-) INCYTE PHARM INC.  
PA  
XX Tang YN, Corley NC, Guegler KJ, Gorgone GA, Patterson C,  
PI Hillman JL, Baughn MR, Lal P, Azimzai Y, Yue H, Yang J;  
XX WPI; 2000-271437/23.  
XX N-PSDB; AAM12401.  
DR  
XX  
XX New polypeptides and polynucleotides, useful for preventing and treating  
PT a disorder associated with increased or decreased expression of RNA  
PT associated proteins.  
XX  
XX  
XX Claim 1; Page 90-91; 131p; English.  
PS  
XX  
XX The present sequence represents a human RNA-associated protein. The  
CC expression of RNA-associated proteins is closely associated with  
CC reproductive tissues, nervous tissues, cell proliferation including  
CC cancer, inflammation and immune responses, and so they may be used for  
CC diagnosis, treatment or prevention of cell proliferative,  
CC immune/inflammatory disorders, and reproductive disorders. Diseases and  
CC disorders which may be treated include actinic keratosis,  
CC atherosclerosis, arteriosclerosis, bursitis, cirrhosis, hepatitis, mixed  
CC connective tissue disease, myelofibrosis, paroxysmal nocturnal  
CC hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia  
CC and cancers, and trauma  
XX  
XX Sequence 216 AA;  
SQ

Query Match	100.0%;	Score 216;	DB 3;	Length 216;
Best Local Similarity	100.0%;	Pred. No. 4.6e-206;		
Matches 216; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	MSRYYLPENP	SLFVNRNVADDT	TSSEDI	RRFEGRYG	PLVDYVPL	DLPTTRPRGAY	QOED	60
Db	1	MSRYYLPENP	SLFVNRNVADDT	TSSEDI	RRFEGRYG	PLVDYVPL	DLPTTRPRGAY	QOED	60
QY	61	VADADALNHLNR	KWICGRQIEI	QFAQGRK	PNOMKAKE	GRNVYSS	SRDDVDR	YRBSR	120
Db	61	VADADALNHLNR	KWICGRQIEI	QFAQGRK	PNOMKAKE	GRNVYSS	SRDDVDR	YRBSR	120
QY	121	SRSYERRRSRS	FDYNYRRSY	SPNRSR	PLTGP	RRREAI	PTIMDQ	NAAGIP	SVLLTLTQ
Db	121	SRSYERRRSRS	FDYNYRRSY	SPNRSR	PLTGP	RRREAI	PTIMDQ	NAAGIP	SVLLTLTQ
QY	181	ERSESGKRTKEG	QFKRPKGGMKYL	QYECYCN	ITLLV				216
Db	181	ERSESGKRTKEG	QFKRPKGGMKYL	QYECYCN	ITLLV				216

RESULT 2  
AAY70220  
ID AAY70220 standard; protein; 216 AA.

XX	AAV70220;
AC	
XX	06-JUN-2000 (first entry)
DT	
XX	
DE	Human RNA-associated protein-1 (RNAP-1).
XX	
KM	RNA-associated protein; RNAP; human; clone 399781; cytosolic;
KM	immunosuppressive; anti-inflammatory; keratolytic; neuroprotective;
KW	antiartherosclerotic; hepatotropic; antipsoriatic; virucide; anti-HIV;
KW	antiallergic; antirheumatic; antiarthritic; ophthalmological; autoimmune;

KW actinobacillary; cell proliferative disorder; inflammation; cirrhosis;  
KW actinic keratosis; buritis; arteriosclerosis; arthrosclerosis;  
KW hepatitis; myelofibrosis; primary thrombocythemia; psoriasis; cancer  
KW mixed connective tissue disease; MCTD; HIV; warts; Crohn's disease  
KW allergy; rheumatoid arthritis; paratuberc infection.

OS Homo sapiens.

FT	Key	Location/Qualifiers
FT	Modified-site	/note= "Potential Glycosylation site"
FT	Binding-site	12. .83
FT		/label= RNP-1/RNP_FNA_binding_motif
FT	Modified-site	21
FT		/note= "Potential Phosphorylation site"
FT	Modified-site	47
FT		/note= "Potential Phosphorylation site"
FT	Modified-site	107
FT		/note= "Potential Phosphorylation site"
FT	Modified-site	108
FT		/note= "Potential Phosphorylation site"
FT	Modified-site	116
FT		/note= "Potential Phosphorylation site"
FT	Modified-site	129
FT		/note= "Potential Phosphorylation site"
FT	Modified-site	138
FT		/note= "Potential Phosphorylation site"
FT	Modified-site	143
FT		/note= "Potential Phosphorylation site"
FT	Modified-site	150
FT		/note= "Potential Phosphorylation site"
FT	Modified-site	161
FT		/note= "Potential Phosphorylation site"
FT	Modified-site	178
FT		/note= "Potential Phosphorylation site"
FT	Modified-site	185
FT		/note= "Potential Phosphorylation site"

PN WO200011171-A2.

AA		
PD	02-MAR-2000.	99WO-US019361.
XX		
XX	20-AUG-1999;	99WO-US019361.
PF		
XX	21-AUG-1998;	98US-0097550P.
PR	12-JAN-1999;	99US-0115639D.
PR		

PA (INCY-) INCYTE PHARM INC.

DR WPI; 2000-237651/20.

DR N-PSDB; AAZ51250.

Human RNA-associated proteins useful in diagnosing, treating and preventing cell proliferative, autoimmune, inflammatory and infectious disorders.

PS Claim 1; Page 77-78; 123pp; English.

The present amino acid sequence for the human RNA-associated protein-1 (RNAP-1), identified in Incyte clone 399781, derived from pRIN0702 library, is expressed in reproductive, nervous and cardiovascular tissues. It has cytosolic, immunosuppressive, keratolytic, antiinflammatory, antiarteriosclerotic, hepatocytic, vitreous, neuroprotective, antiparasitic, anti-HIV, anti-atherogenic, antiinflammatory, ophthalmological, and antimicrobial activity. RNAP antibodies are useful for diagnosis of diseases associated with altered expression or activity of RNAP. It is used to treat cell proliferative, autoimmune, inflammatory and infectious disorders, like acinic keratosis, burns, arteriosclerosis, atherosclerosis, cataracts,

CC hepatitis, myelofibrosis, mixed connective tissue disease (MCTD),  
 CC psoriasis, primary thrombocytopenia and cancer, HIV, allergies, rheumatoid  
 CC arthritis, uveitis, Crohn's disease, and bacterial, viral and parasitic  
 CC infections  
 XX

SQ Sequence 216 AA;

Query Match 100.0%; Score 216; DB 3; Length 216;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-206;  
 Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNVADDTRESEDLRRFGYGPVIVVYPLDFTYRRRGFAVQFED 60  
 DB 1 MSRYLRPNTSLFVNVADDTRESEDLRRFGYGPVIVVYPLDFTYRRRGFAVQFED 60  
 QY 61 VRDAEDALHNLDRKWIICGRQIEICFAQGDRTKPNOMKAKGRNVYSSSRDYDDYRRSR 120  
 DB 61 VRDAEDALHNLDRKWIICGRQIEICFAQGDRTKPNOMKAKGRNVYSSSRDYDDYRRSR 120  
 QY 121 SRSYERRRSRSRSPDYNNRSYSPPNSRPTGPRRRRAIPMIDQTAAGIPSTVLLTTLQ 180  
 DB 121 SRSYERRRSRSRSPDYNNRSYSPPNSRPTGPRRRRAIPMIDQTAAGIPSTVLLTTLQ 180  
 QY 181 ERSESGKRTKEGQFRPKGKMKVLQYECNTLITLV 216  
 DB 181 ERSESGKRTKEGQFRPKGKMKVLQYECNTLITLV 216

RESULT 3  
 AAY70228  
 ID AAY70228 standard; protein, 183 AA.

XX AAY70228;  
 XX  
 DT 06-JUN-2000 (first entry)  
 XX

Human RNA-associated protein-9 (RNAAP-9).

XX RNA-associated protein; RNAAP; human; clone 934406; cytosolic;  
 XX immunosuppressive; anti-inflammatory; keratolytic; neuroprotective;  
 XX antiarteriosclerotic; hepatotropic; antiposoriatic; vinnicide; anti-HIV;  
 XX antiallergic; antirheumatic; antiarthritic; ophthalmological; autoimmune;  
 XX antimicrobial; cell proliferative disorder; inflammation; cirrhosis;  
 XX actinic keratosis; bursitis; arteriosclerosis; artherosclerosis;  
 XX hepatitis; myelofibrosis; primary thrombocytopenia; psoriasis; cancer;  
 XX mixed connective tissue disease; MCTD; HIV; uveitis; Crohn's disease;  
 XX allergy; rheumatoid arthritis; parasitic infection.

OS Homo sapiens.

XX  
 FH Key Location/Qualifiers  
 FT Modified-site 9

FT Domain /note= "Potential glycosylation site"  
 FT 12..83

FT Modified-site /label= RNA\_recognition\_motif  
 FT 21

FT Region /note= "Potential phosphorylation site"  
 FT 30..81

FT Modified-site /note= "Signature sequence of RNP-1"  
 FT 47

FT Modified-site /note= "Potential phosphorylation site"  
 FT 107

FT Modified-site /note= "Potential phosphorylation site"  
 FT 108

FT Modified-site /note= "Potential phosphorylation site"  
 FT 116

FT Modified-site /note= "Potential phosphorylation site"  
 FT 129

FT Modified-site /note= "Potential phosphorylation site"  
 FT 138

FT Modified-site /note= "Potential phosphorylation site"  
 FT 143

FT Modified-site /note= "Potential phosphorylation site"

FT Modified-site 150  
 FT /note= "Potential phosphorylation site"  
 FT 158  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT 160  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT 166  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT 179  
 FT Modified-site /note= "Potential phosphorylation site"  
 FT 180  
 FT /note= "Potential phosphorylation site"

PN W020001171-A2.

PD 02-MAR-2000.

PF 20-AUG-1999; 99WO-US019361.

PR 21-AUG-1998; 98US-0097550P.

PR 12-JAN-1999; 99US-0115639P.

PA (INCY-) INCYTE PHARM INC.

PI Hillman JL, Yue H, Tang YT, Corley NC, Guegler KJ, Gorgone GA;  
 PI Patterson C, Baughn MR, Ial P, Bandman O, Reddy R, Azimzai Y;  
 PI Shih JL, Yang J, Lu DM;

DR WPI, 2000-237651/20.

DR N-PSDB; AAZ51258.

PT Human RNA-associated proteins useful in diagnosing, treating and  
 PT preventing cell proliferative, autoimmune, inflammatory and infectious  
 PT disorders.

XX Claim 1; Page 86; 123p; English.

XX The present amino acid sequence is the human RNA-associated protein-9  
 CC (RNAAP-9), identified in Incyte clone 934406, derived from CERNO101  
 CC library. It is expressed in reproductive, nervous, cardiovascular and  
 CC haematopoietic/immune tissues. It has cytosolic, immunosuppressive,  
 CC anti-inflammatory, antiarteriosclerotic, hepatotropic, keratolytic,  
 CC neuroprotective, antiposoriatic, anti-HIV, antiallergic, antirheumatic,  
 CC vinnicide, antiarthritic, ophthalmological and antimicrobial activity.  
 CC RNAAP antibodies are useful for diagnosis of diseases associated with  
 CC altered expression or activity of RNAAP. It is used to treat cell  
 CC proliferative, autoimmune, inflammatory and infectious disorders, like  
 CC actinic keratosis, bursitis, arteriosclerosis, artherosclerosis,  
 CC cirrhosis, hepatitis, myelofibrosis, mixed connective tissue disease  
 CC (MCTD), psoriasis, primary thrombocytopenia and cancer, HIV, allergies,  
 CC rheumatoid arthritis, uveitis, Crohn's disease, and bacterial, viral and  
 CC parasitic infections

XX Sequence 183 AA;

Query Match 71.8%; Score 155; DB 3; Length 183;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-145;

Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNVADDTRESEDLRRFGYGPVIVVYPLDFTYRRRGFAVQFED 60

DB 1 MSRYLRPNTSLFVNVADDTRESEDLRRFGYGPVIVVYPLDFTYRRRGFAVQFED 60

QY 61 VRDAEDALHNLDRKWIICGRQIEICFAQGDRTKPNOMKAKGRNVYSSSRDYDDYRRSR 120

DB 61 VRDAEDALHNLDRKWIICGRQIEICFAQGDRTKPNOMKAKGRNVYSSSRDYDDYRRSR 120

QY 121 SRSYERRRSRSRSPDYNNRSYSPPNSRPTGPRRR 155

DB 121 SRSYERRRSRSRSPDYNNRSYSPPNSRPTGPRRR 155

RESULT 4

XX	AAB92806
ID	AAB92806 standard; protein; 183 AA.
AC	AAB92806;
DT	26-JUN-2001 (first entry)
DE	Human protein sequence SEQ ID NO:11320.
XX	
KM	Human; primer; detection; diagnosis; antisense therapy; gene therapy.
OS	Homo sapiens.
PN	EPI074617-A2.
XX	
PD	07-FEB-2001.
XX	
PF	28-JUL-2000; 2000EP-00116126.
XX	
PR	29-JUL-1999; 99JP-00248036.
XX	
PR	27-AUG-1999; 99JP-00300253.
XX	
PR	11-JAN-2000; 2000JP-00118776.
XX	
PR	02-MAY-2000; 2000JP-00183767.
XX	
PR	09-JUN-2000; 2000JP-00241899.
XX	
PA	(HELI-) HELIX RES INST.
XX	
PI	Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,
XX	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR	WPI, 2001-318749/34.
XX	
PT	Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
PS	Claim 8; SEQ ID NO 11320; 2537bp + Sequence listing; English.
CC	The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
CC	(a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
SO	Sequence 183 AA;
Query Match	71.8%; Score 155; DB 4; Length 183;
Best Local Similarity	100.0%; Pred. No. 1.5e-145; Indels 0; Gaps 0;
Matches 155; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Dd	MSRYLRPPTSLSFVNNAVADTRSEDLRRFRFGYGEIVDVVPLDFYTRRPRGFAVVOED 60 1 MSRYLRPPTSLSFVNNAVADTRSEDLRRFRFGYGEIVDVVPLDYTRRBRGFAYVED 60 61 VDAADALAHNLDRKMI CGRLQITIGPAQGDRKT PNMOKAKEGNNVSSSRYYDDYDXYRSR 120

D6	61	VRODEDALHNIDRKMVICGRQLEIQFAGSDRTPTPMKAKEGENVSSSRHYDDYDRPRGR	120
Qy	121	SRSYERRRRSRSRSPDYNYRNSYSPPNRSPTGPPRA	155
Db	121	SRSYERRRRSRSRSPDYNYRNSYSPPNRSPTGPPRA	155
RESULT 5			
ID	AAB93180		
AC	AAB93180	standard; protein, 261 AA.	
CC			
DT	26-JUN-2001	(first entry)	
DE		Human protein sequence SEQ ID NO:12124.	
KW		Human; primer; detection; diagnosis; antisense therapy; gene therapy.	
XX			
OS	Homo sapiens.		
PN	EP1074617-A2.		
PD	07-FEB-2001.		
PF	28-JUL-2000; 2000EP-00116126.		
XX			
ER	29-JUL-1999; 99JP-00248036.		
PR	27-AUG-1999; 99JP-00300253.		
PR	11-JAN-2000; 2000JP-00118776.		
PR	02-MAY-2000; 2000JP-00183767.		
PR	09-JUN-2000; 2000JP-00241899.		
XX			
PA	(HELI-) HELIX RES INST.		
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
XX			
XX	WPI; 2001-318749/34.		
PT		Primer sets for synthesizing polynucleotides, particularly the 5602 full-	
PT		length cDNAs defined in the specification, and for the detection and/or	
PT		diagnosis of the abnormality of the proteins encoded by the full-length	
XX		cDNAs.	
XX			
XX			
PS	Claim 8; SEQ ID NO 12124; 2537bp + Sequence listing; English.		
XX			
CC	The present invention describes primer sets for synthesising 5602 full-		
CC	length cDNAs defined in the specification. Where a primer set comprises:		
CC	(a) an oligo-dT primer and an oligonucleotide complementary to the		
CC	complementary strand of a polynucleotide which comprises one of the 5602		
CC	nucleotide sequences defined in the specification, where the		
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination		
CC	of an oligonucleotide comprising a sequence complementary to the		
CC	complementary strand of a polynucleotide which comprises a 5'-end		
CC	sequence and an oligonucleotide comprising a sequence complementary to a		
CC	polynucleotide which comprises a 3'-end sequence, where the		
CC	oligonucleotide comprises at least 15 nucleotides and the combination of		
CC	the 5'-end sequence/3'-end sequence is selected from those defined in the		
CC	specification. The primer sets can be used in antisense therapy and in		
CC	gene therapy. The primers are useful for synthesising polynucleotides,		
CC	particularly full-length cDNAs. The primers are also useful for the		
CC	detection and/or diagnosis of the abnormality of the proteins encoded by		
CC	the full-length cDNAs. The primers allow obtaining of the full-length		
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and		
CC	AAH13631 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95892		
CC	represent human amino acid sequences; and AAH13629 to AAH13632 represent		
CC	oligonucleotides, all of which are used in the exemplification of the		
CC	present invention		
XX			
XX	Sequence 261 AA;		
XX			

Query Match 67.6%; Score 146; DB 4; Length 261;  
Best Local Similarity 100.0%; Pred. No. 1.8e-136;  
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPPTSLFVNNAVDDTSEDLRRFGRGPIVDYVPLDFTYRPRGFAVQVED 60  
DB 1 MSRYLRPPTSLFVNNAVDDTSEDLRRFGRGPIVDYVPLDFTYRPRGFAVQVED 60  
QY 61 VRDAEDALHNIDRKWICGRQIEIQFAQGRKTPNOMKAKEGNVSSRYDDYDRYRSR 120  
DB 61 VRDAEDALHNIDRKWICGRQIEIQFAQGRKTPNOMKAKEGNVSSRYDDYDRYRSR 120  
QY 121 SRSYRRSRSPRYNRRSYSPRN 146  
DB 121 SRSYRRSRSPRYNRRSYSPRN 146

RESULT 6  
ABG00682  
ID ABG00682 standard; protein, 198 AA.  
AC  
XX ABG00682;  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #673.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
OS Homo sapiens.  
XX  
XX W0200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
PI Dmanac RT, Liu C, Tang YT;  
PI WPI; 2001-639362/73.  
DR N-PSDB; AAS64869.  
XX  
XX

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

Claim 20; SEQ ID NO 31041; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 198 AA;

Query Match 66.7%; Score 144; DB 4; Length 198;  
Best Local Similarity 100.0%; Pred. No. 1.4e-134;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LFVNNAVDDTSEDLRRFGRGPIVDYVPLDFTYRPRGFAVQVEDVDAEDALHNL 71  
DB 12 LFVNNAVDDTSEDLRRFGRGPIVDYVPLDFTYRPRGFAVQVEDVDAEDALHNL 71  
QY 72 DRKWCGRQIEIQFAQGRKTPNOMKAKEGNVSSRYDDYDRYRSRSRYRRRSRS 131  
DB 72 DRKWCGRQIEIQFAQGRKTPNOMKAKEGNVSSRYDDYDRYRSRSRYRRRSRS 131  
QY 132 RSPDYNRRSYSPNRRPTGRPR 155  
DB 132 RSPDYNRRSYSPNRRPTGRPR 155

RESULT 7  
ABG15228  
ID ABG15228 standard; protein, 238 AA.  
XX  
XX ABG15228;  
AC  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #15219.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
OS Homo sapiens.  
XX  
XX W0200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
PI Dmanac RT, Liu C, Tang YT;  
PI WPI; 2001-639362/73.  
DR N-PSDB; AAS79415.  
XX  
XX

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

Claim 20; SEQ ID NO 45587; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The

CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

Sequence 238 AA;

Query Match 66.7%; Score 144; DB 4; Length 238;  
Best Local Similarity 100.0%; Pred. No. 1.6e-134; Indels 0; Gaps 0;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LFVNNVADDTRESEDLRREGRYGPVIVDYVPLDFYTRRRCGFAYVQPEDVADALHNL 71  
DB 49 LFVNNVADDTRESEDLRREGRYGPVIVDYVPLDFYTRRRCGFAYVQPEDVADALHNL 108  
QY 72 DRKMICGQIIEIOFAGQDRKTPNQMKAKEGNNYSSSYDYDYYRRRSRSYERRRSRS 131  
DB 109 DRKMICGQIIEIOFAGQDRKTPNQMKAKEGNNYSSSYDYDYYRRRSRSYERRRSRS 168  
QY 132 RSFDYNNRYSYSPNSRPTGRPRR 155  
DB 169 RSFDYNNRYSYSPNSRPTGRPRR 192

RESULT 8  
ABG00681  
ID ABG00681 standard; protein; 118 AA.  
XX  
AC ABG00681;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #672.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
XX  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Dmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
XX  
DR N-PSDB; AAS64868.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 20; SEQ ID NO 31040; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX sequences. (I) is useful as hybridisation probes, polymerase chain  
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
XX and in recombinant production of (II). The polynucleotides are also used  
XX in diagnostics as expressed sequence tags for identifying expressed  
XX genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

Sequence 118 AA;

Query Match 44.4%; Score 96; DB 4; Length 118;  
Best Local Similarity 100.0%; Pred. No. 4.3e-87; Indels 0; Gaps 0;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRTLRPPNTSLFVRNNVADDTRESEDLRREGRYGPVIVDYVPLDFYTRRRCGFAYVQPED 60  
DB 1 MSRTLRPPNTSLFVRNNVADDTRESEDLRREGRYGPVIVDYVPLDFYTRRRCGFAYVQPED 60  
QY 61 VRDAEDALHNLDRKMICGQIIEIOFAGQDRKTPNQM 96  
DB 61 VRDAEDALHNLDRKMICGQIIEIOFAGQDRKTPNQM 96

RESULT 9  
ABG15227  
ID ABG15227 standard; protein; 130 AA.  
XX  
AC ABG15227;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #15218.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
XX  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Dmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
XX  
DR N-PSDB; AAS79414.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 20; SEQ ID NO 45586; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX sequences. (I) is useful as hybridisation probes, polymerase chain  
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
XX and in recombinant production of (II). The polynucleotides are also used

in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG037 represent novel human diagnostic parent did not appear in the invention. Note: The sequence data for this electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 130 AA;

Query Match 44.4%; Score 96; DB 4; Length 130;  
Best Local Similarity 100.0%; Pred. No. 4.ee-87;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVRNVADTRSDLRREFRGYGVYVYPLDPTTRPRGFAYVQED 60  
DB 7 MSRYLRPNTSLFVRNVADTRSDLRREFRGYGVYVYPLDPTTRPRGFAYVQED 66  
QY 61 VRDAEDALHNDLRKWCGRQIEIQFAGCDRKTQNM 96  
DB 67 VRDAEDALHNDLRKWCGRQIEIQFAGCDRKTQNM 102

# RESULT 10

ID AAG02102 standard; protein; 92 AA.

AC AAG02102;

DT 06-OCT-2000 (first entry)

DE Human secreted protein, SEQ ID NO: 6183.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation; gene therapy; chromosome mapping.

OS Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

PF 21-FEB-2000; 2000EP-00200610.

XX 26-FEB-1999; 99US-0122487P.

XX (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

DR N-PDB; AAC02108.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures.

PS Claim 13; SEQ ID NO 6183; 71bp + Sequence Listing; English.

CC The present sequence is a polypeptide encoded by one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated

CC region (UTR) of the mRNA because they are often obtained from oligo-dT CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA CC sequences derived from the 5' ends of mRNAs and even in those cases where CC longer cDNA sequences have been obtained, the full 5' UTR is rarely CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs CC are also used in diagnostic, forensic, gene therapy and chromosome CC mapping procedures. They are used to obtain upstream regulatory sequences CC and to design expression and secretion vectors

Sequence 92 AA;

Query Match 42.6%; Score 92; DB 3; Length 92;  
Best Local Similarity 100.0%; Pred. No. 3.2e-83;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVRNVADTRSDLRREFRGYGVYVYPLDPTTRPRGFAYVQED 60  
DB 1 MSRYLRPNTSLFVRNVADTRSDLRREFRGYGVYVYPLDPTTRPRGFAYVQED 60  
QY 61 VRDAEDALHNDLRKWCGRQIEIQFAGCDRKT 92  
DB 61 VRDAEDALHNDLRKWCGRQIEIQFAGCDRKT 92

# RESULT 11

ID AAE36184 standard; protein; 261 AA.

AC AAE36184;

DT 26-JUN-2003 (first entry)

DE Human CGPD-15 protein.

XX Human; cell growth; differentiation; cell proliferative disorder; CGPD; KW death; cancer; cirrhosis; hepatitis; prolatias; developmental disorder; KW cataract; anaemia; Cushing's syndrome; neurological disorder; epilepsy; KW Huntington's disease; Parkinson's disease; Crohn's disease; transgenic; KW Goodpasture's syndrome; inflammatory disorder; reproductive disorder; KW rheumatoid arthritis; Grave's disease; dementia; autoimmune disorder; KW infertility; endometriosis; placenta disorder; preeclampsia; diabetes; KW metabolic disorder; tumour; cancer; choriocarcinoma; chronic villitis; KW infertility; obesity; transgenic animal; gene therapy.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Domain 12..83

XX /note="RNA recognition motif"

PN W02002102310-A2.

PD 27-DEC-2002.

XX 12-JUN-2002; 2002MO-US018834.

PF 15-JUN-2001; 2001US-0298617P.

PR 21-JUN-2001; 2001US-0300376P.

PR 29-JUN-2001; 2001US-0301873P.

PR 09-JUL-2001; 2001US-0304053P.

PR 13-JUL-2001; 2001US-0305330P.

PR 13-JUL-2001; 2001US-0305361P.

PR 13-JUL-2001; 2001US-0305370P.

XX (INCY-) INCYTE GENOMICS INC.

XX Yue H, Lu DM, Hafalia Afa, Arvizu CS, Ramkumar J, Tang YT;

PI Khan FA, Greene BD, Richardson TW, Yang J, Ison CH, Warren BA;

PI Elliott VS, Emerling BM, Goryad AE, Lee BA, Griffin JA;

PI Zebajadjan Y, Swarnakar A, Lal PG, Baughn MR, Tran UK, Lee S;

PI Foreythe ID, Au-Young JK, Coleman IM;



DR WPI, 2003-167434/16.  
DR N-PSDB; AAD54748.  
PT Novel isolated human protein associated with cell growth,  
PT differentiation, and death, referred as CGPD, useful for diagnosing,  
PT treating, preventing cancer, cirrhosis, cataract, epilepsy, anemia,  
PT infertility, psoriasis.  
XX  
XX Claim 1; Page 163-164; 181pp; English.  
XX  
XX The invention relates to human proteins associated with cell growth,  
XX differentiation and death, referred as CGPD and nucleic acid molecules  
XX encoding such proteins. Sequences of the invention are useful in the  
XX diagnosis, treatment and prevention of cell proliferative disorders (e.g.  
XX cancer, cirrhosis, hepatitis, psoriasis, etc), developmental disorders  
XX (e.g. cataract, anaemia, Cushing's syndrome, etc), neurological disorders  
XX (e.g. epilepsy, Huntington's disease, Parkinson's disease, dementia etc),  
XX autoimmune/inflammatory disorders (e.g. Goodpasture's syndrome, Grave's  
XX disease, rheumatoid arthritis, Crohn's disease, etc), reproductive  
XX disorders (e.g. infertility, endometriosis, endometrial or ovarian  
XX tumour, cancer of the testis, etc), disorders of the placenta (e.g.  
XX pre-eclampsia, choriocarcinoma, chronic villitis, etc) and metabolic  
XX disorders (e.g. obesity or diabetes). Polynucleotides of the invention  
XX are useful for detecting upstream sequences such as promoters and  
XX regulatory elements, creating knock out or knock in humanised animals or  
XX transgenic animals to model human diseases and somatic or germline gene  
XX therapy for treating the disorders. They may also be used for generating  
XX hybridisation probes useful in mapping the naturally occurring genomic  
XX sequences and for developing genetic linkage maps, detecting differences  
XX in chromosomal location due to translocation, inversion etc. They are  
XX also useful for generating a transcript image of a tissue or cell type.  
XX The present sequence is human CGPD protein  
SQ Sequence 261 AA;

Query Match 12.5%; Score 27; DB 6; Length 261;  
Best Local Similarity 100.0%; Pred. No. 2.8e-18;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 PLDFYTRRRGRFAYVQFEDVRDAEDAL 68  
|||  
DB 42 PLDFYTRRRGRFAYVQFEDVRDAEDAL 68

RESULT 12  
ID AAM41266  
XX AAM41266 standard; protein; 271 AA.  
AC AAM41266;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 6197.

XX Human; nocrotropic; immunosuppressant; cytoskeletal; gene therapy; cancer;  
XX peripheral nervous system; neuropathic; central nervous system; CNS;  
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemocarcinoma;  
XX chemokine; thrombolytic; drug screening; arthritis; inflammation;  
XX leukaemia.

XX Homo sapiens.  
XX WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US034263.

XX 23-DEC-1999; 99US-00471275.  
XX 21-JAN-2000; 2000US-00488725.  
XX 25-APR-2000; 2000US-0052317.  
XX 20-JUN-2000; 2000US-00598042.

PR 19-JUL-2000; 2000US-00620312.  
PR 03-AUG-2000; 2000US-00653450.  
PR 14-SEP-2000; 2000US-00662191.  
PR 19-OCT-2000; 2000US-00693036.  
PR 29-NOV-2000; 2000US-00727344.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
XX Zhou P, Goodrich R, Drmanac RT;  
XX  
XX WPI; 2001-442253/47.  
DR N-PSDB; AAI60422.

XX Novel nucleic acids and polypeptides, useful for treating disorders such  
XX as central nervous system injuries.

Example 2; SEQ ID NO 6197; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and the  
XX encoded polypeptides (AAM38642-AA42213) with nocrotropic,  
XX immunosuppressant and cytoskeletal activity. The polynucleotides are useful  
XX in gene therapy. A composition containing a polypeptide or polynucleotide  
XX of the invention may be used to treat diseases of the peripheral nervous  
XX system, such as peripheral nervous injuries, peripheral neuropathy and  
XX localised neuropathies and central nervous system diseases, such as  
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
XX utilisation of the activities such as: immune system suppression,  
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
XX assays for receptor activity, arthritis and inflammation, leukaemias and  
XX C.N.S disorders. Note: The sequence data for this patent did not form  
XX part of the printed specification

Query Match 12.5%; Score 27; DB 4; Length 271;  
Best Local Similarity 100.0%; Pred. No. 2.9e-18;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 PLDFYTRRRGRFAYVQFEDVRDAEDAL 68  
|||  
DB 52 PLDFYTRRRGRFAYVQFEDVRDAEDAL 78

RESULT 13  
ID ABU11773  
XX ABU11773 standard; protein; 280 AA.  
AC ABU11773;

DT 13-FEB-2003 (first entry)

DE Human MDDT polypeptide SEQ ID 720.

XX MDDT; human; disease detection and treatment molecule polypeptide;  
XX anti-inflammatory; immunosuppressive; osteopathic; cytoskeletal; anti-HIV;  
XX haemostatic; nephrotropic; anti-anaemic; antipsoriatic; hepatotropic;  
XX gene therapy; protein replacement therapy; cell proliferative disorder;  
XX cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; sarcoma;  
XX anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;  
XX Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopenia;  
XX psoriasis; hepatitis.

XX Homo sapiens.  
XX WO200279449-A2.

PD 10-OCT-2002.

PF 27-MAR-2002; 2002WO-US009944.

XX 28-MAR-2001; 2001US-0279619P.  
 PR 29-MAR-2001; 2001US-0280067P.  
 PR 29-MAR-2001; 2001US-0280068P.  
 PR 16-MAY-2001; 2001US-0291280P.  
 PR 17-MAY-2001; 2001US-0291829P.  
 PR 19-JUN-2001; 2001US-0291849P.  
 PR 20-JUN-2001; 2001US-0299422P.  
 PR 20-JUN-2001; 2001US-0299776P.  
 PR 20-JUN-2001; 2001US-0300001P.  
 XX  
 XX (INCYTE GENOMICS INC.)  
 XX  
 PA Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;  
 PI Dufour GE, Hillman JL, Yu YF, Tuason O, Yap PR, Amshey SR;  
 PI Dargabety SC, Dam TC, Liu TF, Nguyen DA, Kleeefeld Y, Gerstin EH;  
 PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;  
 PI Flores V, Marwaha R, Lo A, Lan RY, Urashika ME;  
 XX  
 DR WPI: 2003-058431/05.  
 DR N-PSDB; ABX34763.  
 XX  
 PT New purified disease detection and treatment molecule proteins and  
 PT polynucleotides, useful for diagnosing, treating or preventing cancers  
 PT (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis  
 PT or hepatitis.  
 XX  
 PS Claim 27; SEQ ID NO 720; 339pp + Sequence Listing; English.  
 XX  
 CC This invention describes a novel disease detection and treatment molecule  
 CC polypeptide (MDPT) which has anti-inflammatory, immunosuppressive,  
 CC osteopathic, cyostatic, anti-HIV, haemostatic, nephrotropic,  
 CC antiaiemic, antipsoriatic and hepatotropic activity. The polynucleotides  
 CC and the polypeptides of the invention can be used for gene therapy,  
 CC protein replacement therapy and are useful for treating a variety of  
 CC diseases or conditions. These polypeptides or polynucleotides are  
 CC particularly useful for diagnosing, treating or preventing cell  
 CC proliferative disorders (e.g. cancers including adenocarcinoma,  
 CC leukemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's  
 CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's  
 CC syndromes, inflammation, osteoporosis, thrombocytopenia, psoriasis or  
 CC hepatitis; ABU11450-ABU11845 represent the MDPT polynucleotides encoded  
 CC by ABU11450-ABU11845, described in the disclosure of the invention. NOTE:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 XX  
 SQ Sequence 280 AA:  
 Query Match 12.5%; Score 27; DB 6; Length 280;  
 Best Local Similarity 100.0%; Pred. No. 3e-18;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 42 PLDFYTRRPRGFAVQVEDVDAEDAL 68  
 DB 61 PLDFYTRRPRGFAVQVEDVDAEDAL 87  
 RESULT 14  
 ID AAM39480  
 XX AAM39480 standard; protein; 287 AA.  
 AC AAM39480;  
 DT 22-OCT-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 2625.  
 XX  
 XX Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; cancer;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemorecttic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

KW leukemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 26-DEC-2000; 2000WO-US034263.  
 PF  
 XX  
 PR 23-DEC-1999; 99US-00471275.  
 PR 21-JAN-2000; 2000US-0048725.  
 PR 25-APR-2000; 2000US-00552317.  
 PR 20-JUN-2000; 2000US-00598042.  
 PR 19-JUL-2000; 2000US-00620312.  
 PR 03-AUG-2000; 2000US-00653450.  
 PR 14-SEP-2000; 2000US-00662191.  
 PR 19-OCT-2000; 2000US-00693036.  
 PR 29-NOV-2000; 2000US-00727344.  
 XX  
 XX (HYSEQ- ) HYSEQ INC.  
 PA  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
 PI Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI: 2001-442253/47.  
 DR N-PSDB; AA158636.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders such  
 PT as central nervous system injuries.  
 XX  
 PS Example 4; SEQ ID NO 2625; 10078pp; English.  
 XX  
 CC The invention relates to human nucleic acids (AA157798-AA161369) and the  
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cyostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemorecttic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukemias and  
 CC C.N.S disorders. Note: The sequence data for this patent did not form  
 CC part of the printed specification  
 CC  
 XX  
 SQ Sequence 287 AA:  
 Query Match 12.5%; Score 27; DB 4; Length 287;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-18;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 42 PLDFYTRRPRGFAVQVEDVDAEDAL 68  
 DB 68 PLDFYTRRPRGFAVQVEDVDAEDAL 94  
 RESULT 15  
 ID ABB61853  
 XX ABB61853 standard; protein; 706 AA.  
 AC ABB61853;  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 12351.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.

XX Drosophila melanogaster.  
 OS  
 XX WO200171042-A2.  
 PN  
 XX 27-SEP-2001.  
 PD  
 XX 23-MAR-2001; 2001WO-US009231.  
 PF  
 XX 23-MAR-2000; 2000US-0191637P.  
 PR  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 PI  
 XX WPI; 2001-656860/75.  
 DR  
 DR N-PSDB; ABL05956.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 PT  
 PS Disclosure; SEQ ID NO 12351; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signaling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 706 AA;  
 QY 117 RRSRSRSYERRR 128  
 Db 145 RRSRSRSYERRR 156

Search completed: July 22, 2004, 17:18:50  
 Job time : 56 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 17:17:52 ; Search time 16 Seconds  
(without alignments)  
1298.586 Million cell updates/sec

Title: US-09-787-491B-1  
Perfect score: 216  
Sequence: 1 MSRLRPNTSLFVRNVAD.....FKGGMKYLQYECNILLTLV 216  
Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues  
Word size: 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

PIR 78:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	4.2	229	T12483	hypothetical prote
2	9	4.2	347	T06671	hypothetical prote
3	9	4.2	524	I55595	splicing factor -
4	8	3.7	55	VHNVAC	nucleocapsid prote
5	8	3.7	58	S34045	protamine - North
6	8	3.7	65	VHNVBM	nucleocapsid prote
7	8	3.7	136	D35846	transformer-2 sex-
8	8	3.7	163	I38191	nucleic acid bindi
9	8	3.7	179	T34145	hypothetical prote
10	8	3.7	196	E84638	probable R3p22 sp
11	8	3.7	199	S26047	transformer sex-de
12	8	3.7	201	S26045	sex-determining pr
13	8	3.7	221	B42701	PR264 protein - ch
14	8	3.7	221	A42701	splicing factor SF
15	8	3.7	238	A57198	major polydextral c
16	8	3.7	252	D72866	transformer-2 sex-
17	8	3.7	256	A32373	hypothetical prote
18	8	3.7	264	A31638	RNA-binding protei
19	8	3.7	281	T26084	hypothetical prote
20	8	3.7	286	T09704	hypothetical prote
21	8	3.7	288	S68798	hypothetical prote
22	8	3.7	299	T25596	hypothetical prote
23	8	3.7	312	T26085	hypothetical prote
24	8	3.7	315	T41868	hypothetical prote
25	8	3.7	344	S59043	splicing factor SR
26	8	3.7	345	T33906	hypothetical prote
27	8	3.7	363	B84832	hypothetical prote
28	8	3.7	378	S04336	UI snRNP 70K prote
29	8	3.7	448	A36311	70K UI small nucle

30	8	3.7	471	2	S02016	UI snRNP 70K prote
31	8	3.7	488	2	S36570	E2 protein - human
32	8	3.7	568	2	B96634	probable U2 snRNP
33	8	3.7	614	2	A25707	UI snRNP 70K prote
34	8	3.7	672	2	T21469	hypothetical prote
35	8	3.7	708	2	T25725	hypothetical prote
36	8	3.7	886	2	A32758	beta-amyloid-like
37	8	3.7	891	2	G84693	probable prolins-R
38	8	3.7	919	2	S28179	alpha-amyliase (EC
39	8	3.7	1032	2	A57514	RNA helicase HEH11
40	8	3.7	1173	2	T31421	C-terminal domain-
41	8	3.2	20	2	S28405	lamtin B receptor -
42	7	3.2	61	2	S39425	prolamine P1 duc
43	7	3.2	114	2	C90219	85U ribosomal prot
44	7	3.2	130	2	T20613	hypothetical prote
45	7	3.2	132	2	H70759	hypothetical prote

## ALIGNMENTS

RESULT 1  
T12483  
Hypothetical protein DKFP564B0769.1 - human (Fragment)  
C/Species: Homo sapiens (man)  
C/Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 23-Jul-1999  
C/Accession: T12483  
R/Ansoorge, W.; Wilkner, U.; Mewes, H.W.; Gassenhuber, U.; Wiemann, S.  
submitted to the Protein Sequence Database, June 1999  
A/Reference number: Z17527  
A/Accession: T12483  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-299 <ANS>  
A/Cross-references: EMBL:AL080186  
A/Experimental source: fetal brain; clone DKFP564B0769  
C/Genetics:  
A/Note: DKFP564B0769.1

Query Match  
Best Local Similarity 100.0%; Pred. No. 0.32;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 ERRRSRSRS 133  
DB 109 ERRRSRSRS 117

RESULT 2  
T06671  
hypothetical protein T17F15.10 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 22-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 22-Oct-1999  
C/Accession: T06671  
R/Quetier, F.; Choisme, N.; Robert, C.; Brottier, P.; Mincker, P.; Catolico, L.; Artigau  
submitted to the Protein Sequence Database, April 1999  
A/Reference number: Z15793  
A/Accession: T06671

A/Molecule type: DNA  
A/Residues: 1-347 <QUE>  
A/Cross-references: EMBL:AL049658; GSPDB:GN00061; ATSP:T17F15.10  
C/Genetics:  
A/Experimental source: cultivar Columbia; BAC clone T17F15  
A/Gene: ANSP:T17F15.10  
A/Map position: 3  
A/Introns: 188/3; 206/1; 235/1; 251/2

Query Match  
Best Local Similarity 100.0%; Pred. No. 0.37;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 125 ERRRSRSRS 133  
|||||||

Db 129 ERRRSRSRS 137

RESULT 3

155595  
C:Species: Homo sapiens (man)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 24-Sep-1999  
C:Accession: 155595  
R:Imai, H.; Chan, E.K.; Kiyosawa, K.; Fu, X.D.; Tan, E.M.  
J. Clin. Invest. 92, 2419-2426, 1993  
A:Title: Novel nuclear autoantigen with splicing factor motifs identified with antibody  
A:Reference number: 155595; MUID:94043761; PMID:8227358  
A:Accession: 155595  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-524 <RES>  
A:Cross-references: GB:LI0910; NID:9405191; PIDN:AAA16346.1; PID:9405192  
C:Genetics:  
A:Gene: CCL3  
C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein  
F:154-220/Domain: ribonucleoprotein repeat homology <RMI>  
F:251-318/Domain: ribonucleoprotein repeat homology <RMI>

Query Match 4.2%; Score 9; DB 2; Length 524;  
Best Local Similarity 100.0%; Pred. No. 0.51;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 ERRRSRSRS 133  
Db 78 ERRRSRSRS 86

RESULT 4

VHNVAC  
nucleocapsid protein - Autographa californica nuclear polyhedrosis virus  
N:Alternate names: arginine-rich 6.9k protein  
C:Species: Autographa californica nuclear polyhedrosis virus, AcMNPV  
C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 24-Nov-1999  
C:Accession: A26593; E72862  
R:Wilson, M.E.; Mainprize, T.H.; Friesen, P.D.; Miller, L.K.  
J. Virol. 61, 661-666, 1987  
A:Title: Location, transcription, and sequence of a baculovirus gene encoding a small an  
A:Reference number: A26593; MUID:87112972; PMID:3543402  
A:Accession: A26593  
A:Molecule type: mRNA  
A:Residues: 1-55 <MIL>  
A:Cross-references: GB:M15370; NID:9332483; PIDN:AAA46745.1; PID:9332484  
R:Ayres, M.D.; Howard, S.C.; Kuzio, J.; Lopez-Ferber, M.; Possee, R.D.  
Virology 202, 586-605, 1994  
A:Title: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus.  
A:Reference number: A72850; MUID:94303173; PMID:8030224  
A:Accession: E72862  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-55 <ATR>  
A:Cross-references: GB:LI2858; NID:9510708; PIDN:AAA66730.1; PID:9559169  
C:Genetics:  
A:Gene: Ac-PC.9  
C:Superfamily: Autographa californica nuclear polyhedrosis virus nucleocapsid protein  
C:Keywords: nucleocapsid

Query Match 3.7%; Score 8; DB 1; Length 55;  
Best Local Similarity 100.0%; Pred. No. 0.87;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 YRRRSRSRS 123  
Db 36 YRRRSRSRS 43

RESULT 5  
S34045

protamine - North American opossum

C:Species: Didelphis virginiana, Didelphis marsupialis virginiana (North American opossum  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 23-Jul-1999  
C:Accession: S34045  
R:Winkler, R.J.; Nishikawa, S.; Connor, W.; Dixon, G.H.  
Eur. J. Biochem. 215, 63-72, 1993  
A:Title: Characterization of a marsupial sperm protamine gene and its transcripts from ti  
A:Reference number: S34045; MUID:93345500; PMID:8344286  
A:Accession: S34045  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-58 <MID>  
A:Cross-references: EMBL:X74044; NID:9407062; PIDN:CAA52193.1; PID:9407063  
C:Superfamily: sperm histone  
C:Keywords: DNA binding; nucleus

Query Match 3.7%; Score 8; DB 2; Length 58;  
Best Local Similarity 100.0%; Pred. No. 0.91;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRRSRSRS 133  
Db 5 RRRRSRSRS 12

RESULT 6

VHNVBM  
nucleocapsid protein DNA binding P6.9 orf100 - Bombyx mori nuclear polyhedrosis virus (1;  
N:Alternate names: DNA-binding protein  
C:Species: Bombyx mori nuclear polyhedrosis virus, BMSNPV  
A:Variety: isolate T3  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 11-May-2000  
C:Accession: A38554; T41841  
R:Maeda, S.; Kamita, S.G.; Kataoka, H.  
Virology 180, 807-810, 1991  
A:Title: The basic DNA-binding protein of Bombyx mori nuclear polyhedrosis virus: the ex  
A:Reference number: A38554; MUID:91112001; PMID:1703373  
A:Accession: A38554  
A:Molecule type: DNA  
A:Residues: 1-65 <MAE>  
A:Cross-references: EMBL:M63416; NID:9332494; PIDN:AAA46691.1; PID:9332495  
R:Gomi, S.; Majima, K.; Maeda, S.  
J. Gen. Virol. 80, 1333-1337, 1999  
A:Title: Sequence analysis of the genome of Bombyx mori nucleopolyhedrovirus.  
A:Reference number: Z2020; MUID:99281911; PMID:10355780  
A:Accession: T41841  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-65 <RAM>  
A:Cross-references: EMBL:L33180; PIDN:AA063770.1  
A:Experimental source: isolate T3  
C:Superfamily: Autographa californica nuclear polyhedrosis virus nucleocapsid protein  
C:Keywords: DNA binding; late protein; nucleocapsid

Query Match 3.7%; Score 8; DB 1; Length 65;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 YRRRSRSRS 123  
Db 46 YRRRSRSRS 53

RESULT 7

D35846  
transformer-2 sex-determining protein, form D - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 23-Jul-1999  
C:Accession: D35846  
R:Mattox, W.; Palmer, M.J.; Baker, B.S.  
Genes Dev. 4, 789-805, 1990  
A:Title: Alternative splicing of the sex determination gene transformer-2 is sex-specific  
A:Reference number: A35846; MUID:90337316; PMID:2116360

A:Accession: D35846  
A:Molecule type: DNA  
A:Residues: 1-136 <MAT>  
A:Cross-references: GB:M30939; NID:9158648; PIDN:AAA28957.1; PID:9158652  
C:Genetics:  
A:Gene: tra-2  
A:Cross-references: FlyBase:FBgn0003742  
C:Superfamily: transformer-2 sex-determining protein; ribonucleoprotein repeat homology  
C:Keywords: alternative splicing

Query Match 3.7%; Score 8; DB 2; Length 136;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSS 133  
DB 56 RRRSRSS 63

RESULT 8  
I38191  
nucleic acid binding protein - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 05-Nov-1999  
C:Accession: I38191  
R:Nakamura, Y.; Morinuchi, R.; Nakayama, D.; Yamashita, I.; Higashiyama, Y.; Yamamoto, T.  
J. Gen. Virol. 75, 2625-2633, 1994  
A:Title: Altered expression of a novel cellular gene as a consequence of integration of  
A:Reference number: I38191; MUID:95016514; PMID:7931148  
A:Accession: I38191  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-163 <RES>  
A:Cross-references: EMBL:X76302; NID:9431952; PIDN:CAA53949.1; PID:9431953  
C:Genetics:  
A:Gene: RY-1  
A:Introns: 124/3; 146/2

Query Match 3.7%; Score 8; DB 2; Length 163;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSS 133  
DB 46 RRRSRSS 53

RESULT 9  
T34145  
hypothetical protein C33H5.12 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: T34145  
R:Bradshaw, H.; Stellyes, L.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid C33H5.  
A:Reference number: Z21482  
A:Accession: T34145  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-179 <BRA>  
A:Cross-references: EMBL:U1007; PIDN:AAA82270.1; CESP:C33H5.12  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Gene: CESP:C33H5.12  
A:Introns: 18/3; 62/2; 98/2; 149/2  
C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein

Query Match 3.7%; Score 8; DB 2; Length 179;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 DVDAEDA 67

DB 48 DVDAEDA 55

RESULT 10  
E84638  
probable RSP22 splicing factor [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: E84638  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: E84638  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-196 <STO>  
A:Cross-references: GB:AE002093; NID:94572679; PIDN:AAD23894.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g24590  
A:Map position: 2  
C:Superfamily: glycine-rich RNA-binding protein; ribonucleoprotein repeat homology

Query Match 3.7%; Score 8; DB 2; Length 196;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSS 133  
DB 120 RRRSRSS 127

RESULT 11  
S26047  
transformer sex-determining protein - fruit fly (Drosophila virilis)  
C:Species: Drosophila virilis  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 17-Mar-1999  
C:Accession: S26047  
R:O'Neil, M.T.; Belote, J.M.  
Genetics 131, 113-128, 1992  
A:Title: Interspecific comparison of the transformer gene of drosophila reveals an unu  
A:Reference number: S26044; MUID:92275311; PMID:1552233  
A:Accession: S26047  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-199 <ONE>  
A:Cross-references: EMBL:X66528; NID:99206; PID:99207  
C:Genetics:  
A:Gene: FlyBase:Vtr/tra  
A:Cross-references: FlyBase:FBgn0013147  
A:Introns: 11/3; 138/3

Query Match 3.7%; Score 8; DB 2; Length 199;  
Best Local Similarity 100.0%; Pred. No. 2.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSS 133  
DB 104 RRRSRSS 111

RESULT 12  
S26045  
sex-determining protein transformer - fruit fly (Drosophila hydei)  
C:Species: Drosophila hydei  
C>Date: 19-Jul-1996 #sequence\_revision 19-Jul-1996 #text\_change 21-Jul-2000  
C:Accession: S26045  
R:O'Neill, M.T.; Belote, J.M.  
Genetics 131, 113-128, 1992  
A:Title: Interspecific comparison of the transformer gene of drosophila reveals an unu

A:Reference number: S26044; MUID:92275311; PMID:1592233  
 A:Accession: S26045  
 A:Molecule type: DNA  
 A:Residues: 1-201 <ONE>  
 A:Cross-references: EMBL:X66931; NID:97457; PIDD:CAA47365.1; PID:97458  
 A:Note: the authors did not translate the codon for residue 111  
 C:Genetics:  
 A:Gene: tra  
 A:Cross-references: FlyBase:FBgn0012404  
 A:introns: 11/3; 134/3

Query Match 3.7%; Score 8; DB 2; Length 201;  
 Best Local Similarity 100.0%; Pred. No. 2.7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSRS 133  
 |||||  
 Db 102 RRRSRSRS 109

RESULT 13  
 PR264 protein - chicken  
 C:Species: Gallus gallus (chicken)  
 C:Date: 26-May-1994 #sequence\_revision 26-May-1994 #text\_change 24-Sep-1999  
 C:Accession: B42701; S17327  
 R:Vellard, M.; Sureau, A.; Soret, J.; Martinerie, C.; Perbal, B.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 2511-2515, 1992  
 A:Title: A potential splicing factor is encoded by the opposite strand of the trans-spl  
 A:Reference number: A42701; MUID:92212859; PMID:1557353  
 A:Accession: B42701  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-221 <VEL>  
 A:Cross-references: EMBL:X62446; NID:963751; PIDD:CAA4306.1; PID:963752  
 C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein  
 F:15-82/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match 3.7%; Score 8; DB 2; Length 221;  
 Best Local Similarity 100.0%; Pred. No. 2.7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSRS 133  
 |||||  
 Db 125 RRRSRSRS 132

RESULT 14  
 A42701  
 splicing factor SFRS2 - human  
 N:Alternate names: PR264/SC35; splicing factor arginine/serine rich 2 (SFRS2)  
 C:Species: Homo sapiens (man)  
 C:Date: 26-May-1994 #sequence\_revision 26-May-1994 #text\_change 17-Nov-2000  
 C:Accession: I38128; A42701; A42634; I59216; S17328  
 R:Sureau, A.; Perbal, B.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 932-936, 1994  
 A:Title: Several mRNAs with variable 3' untranslated regions and different stability enc  
 A:Reference number: A36992; MUID:94134745; PMID:8302870  
 A:Accession: I38128  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-221 <SUR>  
 A:Cross-references: EMBL:X75755; NID:9455418; PIDD:CAA53383.1; PID:9455419  
 R:Vellard, M.; Sureau, A.; Soret, J.; Martinerie, C.; Perbal, B.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 2511-2515, 1992  
 A:Title: A potential splicing factor is encoded by the opposite strand of the trans-spl  
 A:Reference number: A42701; MUID:92212859; PMID:1557353  
 A:Accession: A42701  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-221 <VEL>  
 A:Cross-references: EMBL:X62447; NID:935596; PIDD:CAA44307.1; PID:935597  
 R:Pu, X.D.; Maniatis, T.

Science 256, 535-538, 1992  
 A:Title: Isolation of a complementary DNA that encodes the mammalian splicing factor SC3;  
 A:Reference number: A42634; MUID:92237694; PMID:1373910  
 A:Accession: A42634  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-37, 'R', 39-221 <FUA>  
 A:Cross-references: GB:M90104; NID:9337925; PIDD:AAA60306.1; PID:9337926  
 R:Sureau, A.; Soret, J.; Vellard, M.; Crochet, J.; Perbal, B.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 11683-11687, 1992  
 A:Title: The PR264/c-myb connection: expression of a splicing factor modulated by a nucle  
 A:Reference number: I59216; MUID:93101590; PMID:1465383  
 A:Accession: I59216  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-34 <RES>  
 A:Cross-references: GB:L03693; NID:9292398; PIDD:AAA60162.1; PID:9553620  
 C:Genetics:  
 A:Gene: GDB:SFRS2; SC-35  
 A:Cross-references: GDB:I32412; OMIM:600813  
 A:introns: 121/2  
 C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein  
 F:15-82/Domain: ribonucleoprotein repeat homology <RRM1>

Query Match 3.7%; Score 8; DB 2; Length 221;  
 Best Local Similarity 100.0%; Pred. No. 2.7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSRS 133  
 |||||  
 Db 125 RRRSRSRS 132

RESULT 15  
 A57198  
 splicing factor, arginine/serine-rich 7 - human  
 N:Alternate names: splicing factor 9G8  
 C:Species: Homo sapiens (man)  
 C:Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 24-Sep-1999  
 C:Accession: A57198; S46319  
 R:Popelar, M.; Cavaloc, Y.; Mattei, M.G.; Gattoni, R.; Stevenin, J.  
 J. Biol. Chem. 270, 17830-17835, 1995  
 A:Title: The gene encoding human splicing factor 9G8. Structure, chromosome localization  
 A:Reference number: A57198; MUID:95355374; PMID:7629084  
 A:Accession: A57198  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-238 <RES>  
 A:Cross-references: GB:L41887; NID:950423; PIDD:AAA88098.1; PID:950424  
 R:Cavaloc, Y.; Popelar, M.; Fuchs, J.P.; Gattoni, R.; Stevenin, J.  
 EMBO J. 13, 2639-2649, 1994  
 A:Title: Characterization and cloning of the human splicing factor 9G8: a novel 35 kDa f  
 A:Reference number: S46319; MUID:9483389; PMID:8013463  
 A:Accession: S46319  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-238 <CAV>  
 A:Cross-references: EMBL:L22253; NID:9506401; PIDD:AAA3495.1; PID:9506402  
 C:Genetics:  
 A:Gene: GDB:SFRS7; 9G8  
 A:Cross-references: GDB:378350; OMIM:600572  
 A:Map position: 2p22-2p21  
 A:introns: 10/1; 70/2; 129/2; 154/2; 191/2; 209/2; 221/2  
 C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein  
 F:12-74/Domain: ribonucleoprotein repeat homology <RRM3>

Query Match 3.7%; Score 8; DB 2; Length 228;  
 Best Local Similarity 100.0%; Pred. No. 2.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSRS 133  
 |||||  
 Db 125 RRRSRSRS 132



Search completed: July 22, 2004, 17:20:59  
Job time : 16 secs

---

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 17:13:21 ; Search time 13 Seconds

(without alignments)  
865.166 Million cell updates/sec

Title: US-09-787-491B-1

Perfect score: 216  
Sequence: 1 MSRTLRPPNPTSLFVRNVD.....PKGGMKVLQYEXYCTNLTLLV 216

Scoring table: ORIGO  
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	4.2	208	YK69_CAEEL	P34347 caenorhabdi
2	9	4.2	530	RNP2_HUMAN	Q14498 homo sapien
3	9	4.2	530	RNP2_MOUSE	Q6VH51 mus musculu
4	8	3.7	54	BVCP_NPVC	P06545 autographa
5	8	3.7	57	HSP1_DIDMA	P35305 didelphis m
6	8	3.7	58	BVCP_GVCL	P41726 cryptophleb
7	8	3.7	60	HSP1_DASVI	P42135 dasypus vi
8	8	3.7	61	HSP1_SARHA	P42151 sarcophilus
9	8	3.7	64	BVCP_NPYBM	P24649 bombyx mori
10	8	3.7	199	TRSF_DROVI	Q24761 drosophila
11	8	3.7	201	TRSF_DROHY	Q23949 drosophila
12	8	3.7	221	SPR2_CHICK	P03352 gallus gall
13	8	3.7	221	SPR2_HUMAN	O01130 homo sapien
14	8	3.7	221	SPR2_MOUSE	Q62093 mus musculu
15	8	3.7	238	SPR7_HUMAN	Q16629 homo sapien
16	8	3.7	252	VEHE_NPVAC	P24728 autographa
17	8	3.7	264	TR2A_DROME	P19018 drosophila
18	8	3.7	282	TR2A_HUMAN	Q13595 homo sapien
19	8	3.7	288	TR2B_HUMAN	Q13247 homo sapien
20	8	3.7	344	SPR6_HUMAN	Q13247 homo sapien
21	8	3.7	378	RUI7_MOUSE	Q62376 mus musculu
22	8	3.7	437	RUI7_HUMAN	P08611 homo sapien
23	8	3.7	448	RUI7_DROME	P17133 drosophila
24	8	3.7	471	RUI7_XENILA	P09406 xenopus lae
25	8	3.7	488	VE2_HPV49	P36795 human papil
26	8	3.7	653	SPRG_MOUSE	O8CF27 mus musculu
27	8	3.7	659	SPRG_HUMAN	Q6ZM86 homo sapien
28	8	3.7	887	A4_DROME	P14599 drosophila
29	8	3.7	919	AMY_STRLI	Q05884 streptomyce
30	7	3.2	60	HSP1_ORNAN	P35307 ornithochym
31	7	3.2	114	RS17_SULO	O9UX88 sulfolobus
32	7	3.2	126	RS12_CANTP	O8KTP6 candidarus
33	7	3.2	132	YK10_MYCTU	Q10847 mycobacteri

34	7	3.2	148	1	H1L MYTCA	P22974 mytilus cal
35	7	3.2	196	1	SPR2_CAEEL	O09511 caenorhabdi
36	7	3.2	208	1	YKX2_CAEEL	Q10021 caenorhabdi
37	7	3.2	240	1	YD37_MYCTU	Q10647 mycobacteri
38	7	3.2	244	1	YK51_CAEEL	P34433 caenorhabdi
39	7	3.2	303	1	SPR1_ARATH	O22315 arabidopsis
40	7	3.2	306	1	UL31_HSV1	P10215 herpes simp
41	7	3.2	309	1	CYP9_CAEEL	O09637 caenorhabdi
42	7	3.2	325	1	LDH_TACCA	P00343 lactocacilli
43	7	3.2	333	1	SWT_PASMO	P57956 pasteurella
44	7	3.2	339	1	SRRI_HUMAN	Q9UH36 homo sapien
45	7	3.2	375	1	SR55_DROME	P26686 drosophila

## ALIGNMENTS

RESULT 1  
YK69\_CAEEL STANDARD; PRT; 208 AA.  
ID AC P34347:

DT 01-FEB-1994 (Rel. 28, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein C29E4.9 in chromosome III.  
GN C29E4.9.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;

RM [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ranscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,  
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
RA Wohldman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38 (1994).

RM [2]  
RP REVISIONS.  
RA Waterston R.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; I23651; AAA27956.2; -;  
DR PIR; S44765; S44765.  
DR Wormpep; C29E4.9; CE29199.  
KW Hypothetical protein.  
SQ SEQUENCE 208 AA; 23729 MW; FDA57F6609EEB32 CRC64;

Query Match 4.2%; Score 9; DB 1; Length 208;

Best local Similarity 100.0%; Pred. No. 0.07; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 ERRRRSRSS 133  
|||||||

DB 153 ERRSRSSRS 161

RESULT 2  
RNP2\_HUMAN STANDARD; PRT; 530 AA.

AC Q14498; Q14499; (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE RNA-binding region containing protein 2 (Hepatocellular carcinoma protein 1) (splicing factor HCC1).  
GN RNP2 OR HCC1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND CHARACTERIZATION.  
RC TISSUE=Liver;  
RX MEDLINE=94043761; PubMed=8227358;  
RA Imai H., Chan E.K.L., Kiyosawa K., Fu X.-D., Tan E.M.;  
RA "Novel nuclear autoantigen with splicing factor motifs identified with  
RT antibody from hepatocellular carcinoma.";  
RL J. Clin. Invest. 92:2419-2426(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21638749; PubMed=11780052;  
RA Deloukas P., Matthews L.H., Ashurst J., Burton A., Gilbert J.G.R.,  
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
RA Bailey J., Barlow K.P., Bates K.N., Beard L.M., Beare D.M.,  
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
RA Buckley D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Cordy N.R.,  
RA Coulson A.G., Coville G.J., Deadman R., Dhani P.D., Dunn M.,  
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
RA Graham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,  
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.D., Howden P.J.,  
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
RA Levasaiho M.H., Leverha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
RA Marsh V.L., Martin S.L., McCormack L.J., McElay K., McMurtry A.A.,  
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.T.,  
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,  
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,  
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.,  
RA Wilming L., Wray P.W., Hubbard T., Dudbin R.M., Bentley D.R., Beck S.,  
RA Rogers J.;  
RT "The DNA sequence and comparative analysis of human chromosome 20.";  
RL Nature 414:865-871 (2001).

-1- FUNCTION: Transcriptional coactivator for steroid nuclear receptors ESRI/ER-alpha and ESR2/ER-beta, and JUN/AP-1 (By similarity). May be involved in pre-mRNA splicing process.  
-1- SUBUNIT: Interacts with NCOA6 (By similarity).  
-1- SUBCELLULAR LOCATION: Nuclear; concentrated in nuclear speckles. Colocalizes with the core spliceosomal snRNP proteins.  
-1- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=2;  
Name=1; Synonyms=HCC1.4;  
IsoId=Q14498-1; Sequence=Displayed;  
Name=2; Synonyms=HCC1.3;  
IsoId=Q14498-2; Sequence=VSP\_005820;  
-1- TISSUE SPECIFICITY: Widely expressed. Highly expressed in pancreas, skeletal muscle, lung and brain. Expressed at intermediate level in kidney, liver and heart.  
-1- MISCELLANEOUS: Antibodies against RNP2 are present in sera from a patient with hepatocellular carcinoma who developed several autoantibodies.

CC -1- SIMILARITY: Belongs to the splicing factor SR family.  
CC -1- SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L10910; AAA16346.1; -;  
DR EMBL; L10911; AAA16347.1; -;  
DR EMBL; AL357374; CAC11118.1; -;  
DR EMBL; AL357374; CAC11119.1; -;  
DR PIR; I55595; I55595.  
DR HSSP; P11940; ICVT.  
DR Genew; HGNC:15923; RNP2.  
DR GK; Q14498; -;  
DR MIM; 604739; -;  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF00076; Rrm; 3.  
DR SMART; SM00360; RRM; 3.  
DR PROSITE; PS50102; RRM; 1;  
DR PROSITE; PS00030; RRM\_RNP\_1; 2.  
DR Transcription regulation; Activator; Nuclear protein; RNA-binding;  
KM mRNA processing; mRNA splicing; Repeat; Alternative splicing;  
KM Polymorphism.  
KM POLYMERPHISM.  
FT DOMAIN 41 90 ARG/SER-RICH (RS DOMAIN).  
FT DOMAIN 153 230 RNA-BINDING (RRM) 1.  
FT DOMAIN 250 328 RNA-BINDING (RRM) 2.  
FT DOMAIN 445 508 RNA-BINDING (RRM) 3.  
FT DOMAIN 291 355 ACTIVATING DOMAIN (BY SIMILARITY).  
FT DOMAIN 291 406 INTERACTION WITH JUN (BY SIMILARITY).  
FT DOMAIN 355 406 INTERACTION WITH ESRI AND ESR2  
(BY SIMILARITY).  
FT DOMAIN 406 530 INTERACTION WITH NCOA6 (BY SIMILARITY).  
FT DOMAIN 409 416 POLY-ALA.  
FT VARSPPLIC 392 397 Missing (in isoform 2).  
FT VARIANT 2 2 /FTId=VSP\_005820.  
FT /FTId=VAR\_015228.  
FT /FTId=VAR\_015228.  
SQ SEQUENCE 530 AA; 59379 MW; 0CC610356D4AA040 CRC64;  
Query Match 4.2%; Score 9; DB 1; Length 530;  
Best Local Similarity 100.0%; Pred. No. 0.17; Mismatches 0; Gaps 0;  
Matches 9; Conservative 0; Indels 0; Gaps 0;  
DB 125 ERRSRSSRS 133  
DB 78 ERRSRSSRS 86

RESULT 3  
RNP2\_MOUSE STANDARD; PRT; 530 AA.

AC O6VH51; O6VH51; (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE RNA-binding region containing protein 2 (Coactivator of activating protein-1 and estrogen receptors) (Coactivator of AP-1 and ERS)  
GN RNP2 OR CAPER.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), FUNCTION, AND INTERACTION WITH  
RP NCOA6; ESRI; ESR2 AND JUN.  
RC TISSUE=Liver;  
RX MEDLINE=21638469; PubMed=11704680;

RA Jung D.-J., Na S.-Y., Na D.S., Lee J.W.;  
 RT "Molecular cloning and characterization of CAPER, a novel coactivator  
 of activating protein-1 and estrogen receptors.";  
 RL J. Biol. Chem. 277:1229-1234(2002).  
 RN (2)  
 RP SEQUENCE FROM N.A. (ISOFORM 3), AND SEQUENCE OF 102-530 FROM N.A.  
 RP (ISOFORM 1).  
 RC STRAIN=C57BL/6; TISSUE=Mammary fibroblast, and Retina;  
 RX MEDLINE=22389257; PubMed=12477932;  
 RA Strassberg R.U., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner K.H., Schaefer C.F., Shat N.K.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Shat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carrincci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abraham R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Raley J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Buterfield V.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- FUNCTION: Transcriptional coactivator for steroid nuclear  
 receptors ESR1/ER-alpha and ESR2/ER-beta, and JUN/AP-1. May be  
 involved in pre-mRNA splicing process.  
 CC -1- SUBUNIT: Interacts with NCOA6.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- ALTERNATIVE PRODUCTS:  
 Event=Alternative splicing; Named isoforms=3;  
 Comment=Experimental confirmation may be lacking for some  
 isoforms;  
 CC Name=1; Synonyms=HCCL1.4;  
 CC IsoId=Q8VH51.1; Sequence=Displayed;  
 CC Name=2; Synonyms=HCCL1.3;  
 CC IsoId=Q8VH51.2; Sequence=VSP\_005822;  
 CC Name=3;  
 CC IsoId=Q8VH51.3; Sequence=VSP\_005821, VSP\_005822;  
 CC -1- SIMILARITY: Belongs to the splicing factor SR family.  
 CC -1- SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AY061882; AA32373.1; -;  
 DR EMBL; BC004000; AA04000.1; -;  
 DR EMBL; BC030493; AA030493.1; -;  
 DR MGI; MGI:2157953; Rnpg2.  
 DR GO; GO:000515; F:Protein binding; IPI.  
 DR InterPro; IPR006509; CCL\_SF.  
 DR InterPro; IPR005054; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 3.  
 DR SMART; SM00360; RRM; 2.  
 DR TIGRfam; TIGR01622; SF-CC1; 1.  
 DR PROSITE; PSS0102; RRM; 2.  
 DR PROSITE; PSS0030; RRM\_RNP\_1; 2.  
 KW Transcription regulation; Activator; Nuclear protein; RNA-binding;  
 KM mRNA processing; mRNA splicing; Repeat; Alternative splicing.  
 FT DOMAIN 41 90  
 FT DOMAIN 153 230 RNA-BINDING (RRM) 1.  
 FT DOMAIN 250 328 RNA-BINDING (RRM) 2.  
 FT DOMAIN 445 508 RNA-BINDING (RRM) 3.

FT DOMAIN 291 355 ACTIVATING DOMAIN.  
 FT DOMAIN 291 406 INTERACTION WITH JUN.  
 FT DOMAIN 355 406 INTERACTION WITH ESR1 AND ESR2.  
 FT DOMAIN 406 530 INTERACTION WITH NCOA6.  
 FT DOMAIN 409 416 POLY-ALA.  
 FT VASPLIC 1 157 Missing (in isoform 3).  
 FT VASPLIC 392 397 Missing (in isoform 2 and isoform 3).  
 FT VASPLIC 194 194 /FtId=VSP\_005822.  
 FT CONFLICT 208 208 R -> K (IN REF. 2).  
 FT CONFLICT 208 208 R -> P (IN REF. 2).  
 SQ SEQUENCE 530 AA; 59493 MW; 3BB891A8B1A8CDE CRC64;  
 Query Match 4.2%; Score 9; DB 1; Length 530;  
 Best Local Similarity 100.0%; Pred. No. 0.17;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 125 ERRSRSSRS 133  
 Db 78 ERRSRSSRS 86  
 RESULT 4  
 BVCP NPVAC STANDARD; PRT; 54 AA.  
 ID BVCP NPVAC  
 AC P06545;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE DNA-binding protein (Arginine-rich protein 6.9 kDa) (Basic viral core  
 protein) (Nucleocapsid protein).  
 GN P6.9.  
 OS Autographa californica nuclear polyhedrosis virus (AcMNPV).  
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae.  
 CC Nucleopolyhedrovirus.  
 OX NCBI\_TaxId=46015;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LA;  
 RX MEDLINE=87112972; PubMed=3543402.  
 RA Wilson M.E., Mainprize T.H., Friesen P.D., Miller L.K.;  
 RT "Location, transcription, and sequence of a baculovirus gene encoding  
 a small arginine-rich polypeptide.";  
 RL J. Virol. 61:661-666(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C6;  
 RX MEDLINE=94303173; PubMed=8030224;  
 RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;  
 RT "The complete DNA sequence of Autographa californica nuclear  
 polyhedrosis virus.";  
 RL Virology 202:586-605(1994).  
 RN [3]  
 RP SEQUENCE OF 11-54 FROM N.A.  
 RC STRAIN=HR3;  
 RX MEDLINE=91134998; PubMed=1994581;  
 RA Lu A., Carstens E.B.;  
 RT "Nucleotide sequence of a gene essential for viral DNA replication in  
 the baculovirus Autographa californica nuclear polyhedrosis virus.";  
 RL Virology 181:336-347(1991).  
 RN [4]  
 RP SEQUENCE OF 1-30 FROM N.A.  
 RC STRAIN=HR3;  
 RA Lu A., Craig A., Carstens E.B.;  
 RT Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Thought to be responsible for DNA condensation during  
 packaging of the nucleocapsids.  
 CC -1- PTM: Probably phosphorylated in infected cells.  
 CC -1- SIMILARITY: STRONG TO ARGININE RICH DNA-BINDING PROTEIN OF OTHER  
 BACULOVIRUSES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M15370; AAA46745.1; -  
 DR EMBL; M57687; AAA67905.1; -  
 DR EMBL; L22858; AAA66730.1; -  
 DR EMBL; U10885; AAB08766.1; -  
 DR PIR; A26593; VHWAVC.  
 KW Nucleocapsid; DNA-binding; Late protein; Repeat; Phosphorylation.  
 FT INT MET 0 0 BY SIMILARITY  
 FT DOMAIN 4 23 2 X 6 AA REPEATS OF R-R-R-S-S.  
 FT REPEAT 4 9 1.  
 FT REPEAT 18 23 2.  
 SQ SEQUENCE 54 AA; 6753 MW; 4FBB136721F1AC35 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 54;  
 Best Local Similarity 100.0%; Pred. No. 0.24;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 YRRSRSS 123  
 DB 35 YRRSRSS 42

RESULT 5  
 HSP1\_D1DMA STANDARD; PRT; 57 AA.  
 ID\_HSP1\_D1DMA  
 AC P35305;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Sperm protamine P1.  
 GN PRM1.  
 OS Didelphis marsupialis virginiana (North American opossum), and  
 OS Monodelphis domestica (short-tailed grey opossum).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Metatheria; Didelphimorphia; Didelphidae; Didelphis.  
 OX NCBI\_TaxID=9267, 13616;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=D.marsupialis;  
 RX MEDLINE=93345500; PubMed=8344286;  
 RA Winkfein R.J., Nishikawa S., Connor W., Dixon G.H.;  
 RT "Characterization of a marsupial sperm protamine gene and its  
 RT transcripts from the North American opossum (Didelphis  
 RT marsupialis).";  
 RL Eur. J. Biochem. 215:63-72(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.domestica;  
 RX MEDLINE=95215351; PubMed=7700877;  
 RA Relief J.D., Krajewski C., Westerman M., Winkfein R.J., Dixon G.H.;  
 RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";  
 RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).  
 CC -1- FUNCTION: Protamines substitute for histones in the chromatin of  
 CC sperm during the haploid phase of spermatogenesis. They compact  
 CC sperm DNA into a highly condensed, stable and inactive complex.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: Testis.  
 CC -1- SIMILARITY: Belongs to the protamine P1 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L17007; AAA02812.1; -

DR EMBL; X74044; CAA52193.1; -  
 DR EMBL; L35448; AAA74612.1; -  
 DR PIR; S34045; S34045.  
 DR InterPro; IPR000221; Protamine\_P1.  
 DR Pfam; PF00260; Protamine\_P1; 1.  
 DR PROSITE; PS00048; PROTAmine\_P1; 1.  
 KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
 KW Testis; DNA condensation; Nuclear protein.  
 FT INT MET 0 0 BY SIMILARITY  
 SQ SEQUENCE 57 AA; 7810 MW; 263715B280214E52 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 57;  
 Best Local Similarity 100.0%; Pred. No. 0.25;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSS 133  
 DB 4 RRRSRSS 11

RESULT 6  
 BVCP\_GVCL STANDARD; PRT; 58 AA.  
 ID\_BVCP\_GVCL  
 AC P41726;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE DNA-binding protein (Arginine-rich protein 7.3 kDa) (Basic viral core  
 DE protein) (Nucleocapsid protein).  
 GN P7.3.  
 OS Cryptophlebia leucocretia granulosis virus (C1GV) (Cryptophlebia  
 OS leucocretia granulovirus).  
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae; Granulovirus.  
 OX NCBI\_TaxID=35254;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV3;  
 RX MEDLINE=94292936; PubMed=8021613;  
 RA Jehle J.A., Backhaus H.;  
 RT "Genome organization of the DNA-binding protein gene region of  
 RT Cryptophlebia leucocretia granulosis virus is closely related to that  
 RT of nuclear polyhedrosis viruses.";  
 RL J. Gen. Virol. 75:1815-1820(1994).  
 CC -1- FUNCTION: Thought to be responsible for DNA condensation during  
 CC packaging of the nucleocapsids.  
 CC -1- PTM: Probably phosphorylated in infected cells.  
 CC -1- SIMILARITY: STRONG TO ARGININE RICH DNA-BINDING PROTEIN OF OTHER  
 CC BACULOVIRUSES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X77048; CAA54337.1; -  
 KW Nucleocapsid; DNA-binding; Late protein; Repeat; Phosphorylation.  
 SQ SEQUENCE 58 AA; 7291 MW; 279387A5590C2E CRC64;

Query Match 3.7%; Score 8; DB 1; Length 58;  
 Best Local Similarity 100.0%; Pred. No. 0.26;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSS 133  
 DB 12 RRRSRSS 19

RESULT 7  
 HSP1\_DASVI STANDARD; PRT; 60 AA.  
 ID\_HSP1\_DASVI





BACULOVIRUSES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M63416; AAA46691.1; -  
 DR EMBL: L33180; AAC63770.1; -  
 DR PIR: A38554; VHWYBM.  
 KW Nucleocapsid; DNA-binding; Late protein; Repeat; Phosphorylation.  
 FT INT MET 0 0  
 FT DOMAIN 4 33 3 X 6 AA REPEATS OF R-R-R-R-S-S.  
 FT REPEAT 4 9 1.  
 FT REPEAT 18 23 2.  
 FT REPEAT 28 33 3.  
 FT REPEAT 64 AA; 7947 MW; 0F1BD3BD1C68B9AE CRC64;  
 SQ SEQUENCE

Query Match 3.7%; Score 8; DB 1; Length 64;  
 Best Local Similarity 100.0%; Pred. No. 0.28;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 116 YRRSRSSRS 123  
 DB 45 YRRSRSSRS 52

RESULT 10  
 TRSF\_DROVI STANDARD; PRT; 199 AA.  
 ID TRSF\_DROVI  
 AC Q24761;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Female-specific transformer protein.  
 GN TRA.  
 OS Drosophila virilis (Fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 CC NCBI\_TaxID=7244;  
 [1]  
 RP SEQUENCE FROM N.A., AND FUNCTION.  
 RX MEDLINE=92275311; PubMed=1592233;  
 RA O'Neil M.T., Belote J.M.;  
 RT "Interspecific comparison of the transformer gene of Drosophila  
 RT reveals an unusually high degree of evolutionary divergence.";  
 RL Genetics 131:113-128(1992).  
 CC -I- FUNCTION: Member of the regulatory pathway controlling female  
 CC somatic sexual differentiation, regulated by Sxl. Activates  
 CC dsx female-specific splicing by promoting the formation of a  
 CC splicing enhancer complex which consists of tra, tra2 and sr  
 CC proteins.  
 CC -I- SUBCELLULAR LOCATION: Nuclear; speckled subnuclear compartment (by  
 CC similarity).  
 CC -I- DOMAIN: RS domain directs localization of proteins to the speckled  
 CC subnuclear compartment and the purpose of this localization is to  
 CC allow colocalization and co-concentration of components of the  
 CC splicing and splicing regulatory machinery to permit relatively  
 CC high rates and/or efficiencies of reaction and interaction (by  
 CC similarity).  
 CC -I- MISCELLANEOUS: The sexual regulation of tra occurs through a  
 CC mechanism of sex-specific alternative RNA splicing. The non-sex-  
 CC specific RNA expressed in males is not translated.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X66931; CAA47365.1; -  
 DR PIR: S26045; S26045.  
 DR FlyBase: FBgn0012404; DmYd\tra.  
 KW Nuclear protein; Sexual differentiation.  
 FT DOMAIN 61 114 ARG/SER-RICH (RS DOMAIN).  
 FT SEQUENCE 201 AA; 23743 MW; 1EB759B2AC0BDB34D CRC64;  
 SQ SEQUENCE

Query Match 3.7%; Score 8; DB 1; Length 201;

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X66528; CAA47141.1; -  
 DR PIR: S26047; S26047.  
 DR FlyBase: FBgn0013147; DmYd\tra.  
 KW Nuclear protein; Sexual differentiation.  
 FT DOMAIN 56 118 ARG/SER-RICH (RS DOMAIN).  
 FT SEQUENCE 199 AA; 23809 MW; 23655867BB539733 CRC64;  
 SQ SEQUENCE

Query Match 3.7%; Score 8; DB 1; Length 199;  
 Best Local Similarity 100.0%; Pred. No. 0.8;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 RRRSRSSRS 133  
 DB 104 RRRSRSSRS 111

RESULT 11  
 TRSF\_DROHY STANDARD; PRT; 201 AA.  
 ID TRSF\_DROHY  
 AC Q23949;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Female-specific transformer protein.  
 GN TRA.  
 OS Drosophila hydei (Fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 CC NCBI\_TaxID=7224;  
 [1]  
 RP SEQUENCE FROM N.A., AND FUNCTION.  
 RX MEDLINE=92275311; PubMed=1592233;  
 RA O'Neil M.T., Belote J.M.;  
 RT "Interspecific comparison of the transformer gene of Drosophila  
 RT reveals an unusually high degree of evolutionary divergence.";  
 RL Genetics 131:113-128(1992).  
 CC -I- FUNCTION: Member of the regulatory pathway controlling female  
 CC somatic sexual differentiation, regulated by Sxl. Activates  
 CC dsx female-specific splicing by promoting the formation of a  
 CC splicing enhancer complex which consists of tra, tra2 and sr  
 CC proteins.  
 CC -I- SUBCELLULAR LOCATION: Nuclear; speckled subnuclear compartment (by  
 CC similarity).  
 CC -I- DOMAIN: RS domain directs localization of proteins to the speckled  
 CC subnuclear compartment and the purpose of this localization is to  
 CC allow colocalization and co-concentration of components of the  
 CC splicing and splicing regulatory machinery to permit relatively  
 CC high rates and/or efficiencies of reaction and interaction (by  
 CC similarity).  
 CC -I- MISCELLANEOUS: The sexual regulation of tra occurs through a  
 CC mechanism of sex-specific alternative RNA splicing. The non-sex-  
 CC specific RNA expressed in males is not translated.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X66931; CAA47365.1; -  
 DR PIR: S26045; S26045.  
 DR FlyBase: FBgn0012404; DmYd\tra.  
 KW Nuclear protein; Sexual differentiation.  
 FT DOMAIN 61 114 ARG/SER-RICH (RS DOMAIN).  
 FT SEQUENCE 201 AA; 23743 MW; 1EB759B2AC0BDB34D CRC64;  
 SQ SEQUENCE

Query Match 3.7%; Score 8; DB 1; Length 201;

Best Local Similarity 100.0%; Pred. No. 0.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSS 133  
Db 102 RRRSRSS 109

## RESULT 12

SFR2\_CHICK STANDARD; PRT; 221 AA.  
ID SFR2\_CHICK  
AC P30352;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Splicing factor, arginine/serine-rich 2 (Splicing factor SC35) (SC-35)  
GN SFRS2.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Thymus;  
RX MEDLINE=92212859; PubMed=157353;  
RA Vellard M., Sureau A., Soret J., Martinerie C., Perbal B.;  
RT "A potential splicing factor is encoded by the opposite strand of the  
trans-spliced c-myc exon";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:2511-2515(1992).  
CC -1- FUNCTION: Necessary for the splicing of pre-mRNA. It is required  
for formation of the earliest ATP-dependent splicing complex and  
interacts with spliceosomal components bound to both the 5' and 3'  
splice sites during spliceosome assembly. It also is required for  
CC -1- ATP-dependent interactions of both U1 and U2 snRNPs with pre-mRNA.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- PTM: Extensively phosphorylated on serine residues in the RS  
domain (by similarity).  
CC -1- SIMILARITY: Belongs to the splicing factor SR family.  
CC -1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X62446; CA44306.1; -;  
DR F1R; B42701; B42701.  
DR HSP; P19339; ISXL.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF00076; rrm; 1.  
DR SMART; SMD00360; RRM; 1.  
DR PROSITE; PSS0102; RRM; 1.  
DR PROSITE; PSS00030; RRM\_RNP\_1; 1.  
KW Nuclear protein; RNA-binding; mRNA splicing; Phosphorylation.  
FT DOMAIN 14 92 RNA-BINDING (RRM).  
FT DOMAIN 111 116 GUY-RICH (HINGE REGION).  
FT DOMAIN 117 221 ARG/SER-RICH (RS DOMAIN).  
SQ SEQUENCE 221 AA; 25524 MW; 75A4D8FP9170F1BF CRC64;

Query Match 3.7%; Score 8; DB 1; Length 221;  
Best Local Similarity 100.0%; Pred. No. 0.88;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSS 133  
Db 125 RRRSRSS 132

RESULT 13  
SFR2\_HUMAN STANDARD; PRT; 221 AA.  
ID SFR2\_HUMAN  
AC O01130;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Splicing factor, arginine/serine-rich 2 (Splicing factor SC35) (SC-35)  
GN SFRS2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. AND SEQUENCE OF 67-84.  
RX MEDLINE=92237694; PubMed=1373910;  
RA Fu X.-D., Maniatis T.;  
RT "Isolation of a complementary DNA that encodes the mammalian splicing  
factor SC35";  
RL Science 256:535-538(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Bone marrow;  
RX MEDLINE=92212859; PubMed=157353;  
RA Vellard M., Sureau A., Soret J., Martinerie C., Perbal B.;  
RT "A potential splicing factor is encoded by the opposite strand of the  
trans-spliced c-myc exon";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:2511-2515(1992).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94134745; PubMed=8302870;  
RA Sureau A., Perbal B.;  
RT "Several mRNAs with variable 3' untranslated regions and different  
stability encode the human PR264/SC35 splicing factor";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:932-936(1994).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=22368257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Sleptchenko M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.U., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buttefield Y.S.N., Krzywinski M.I., Skalek U., Smellus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [5]  
RP SEQUENCE OF 1-34 FROM N.A.  
RX MEDLINE=93101590; PubMed=1465383;  
RA Sureau A., Soret J., Vellard M., Crochet J., Perbal B.;  
RT "The PR264/c-myc connection: expression of a splicing factor  
modulated by a nuclear protooncogene";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:11683-11687(1992).  
RN [6]  
RP SEQUENCE OF 67-83.  
RX MEDLINE=92249775; PubMed=1577227;  
RA Zahler A.M., Lane W.S., Stolk J.A., Roth M.B.;  
RT "SR proteins: a conserved family of pre-mRNA splicing factors";  
RL Genes Dev. 6:837-847(1992).

[7] INTERACTIONS IN SPliceosome ASSEMBLY.  
 RA MEDLINE=94084782; PubMed=8261509;  
 RA Wu J.Y., Maniatis T.;  
 RT "Specific interactions between proteins implicated in splice site  
 RT selection and regulated alternative splicing.";  
 RL Cell 75:1061-1070(1993).  
 RN [8]  
 RP SPECIFICITY FOR BETA-GLOBIN MRNA.  
 RX MEDLINE=93368668; PubMed=8361546;  
 RA Fu X.-D.;  
 RT "Specific commitment of different pre-mRNAs to splicing by single SR  
 RT proteins.";  
 RL Nature 365:82-85(1993).  
 RN [9]  
 RP BINDING TO U1-70K.  
 RX MEDLINE=94187841; PubMed=8139654;  
 RA Kohltz J.D., Jamison S.F., Will C.L., Zuo P., Lührmann R.;  
 RT "The human splicing factors ASF/SF2 and SC35 possess distinct,  
 RT functionally significant RNA binding specificities.";  
 RL EMBO J. 14:3540-3551(1995).  
 RN [11]  
 RP INTERACTION WITH SFRS12.  
 RX MEDLINE=20221551; PubMed=10757789;  
 RA Barnard D.C., Patton J.G.;  
 RT "Identification and characterization of a novel serine-arginine-rich  
 RT splicing regulatory protein.";  
 RL Mol. Cell. Biol. 20:3049-3057(2000).  
 CC -1- FUNCTION: Necessary for the splicing of pre-mRNA. It is required  
 CC for formation of the earliest ATP-dependent splicing complex and  
 CC interacts with spliceosomal components bound to both the 5' and 3'  
 CC splice sites during spliceosome assembly. It also is required for  
 CC ATP-dependent interactions of both U1 and U2 snRNPs with pre-  
 CC mRNA. Interacts with other spliceosomal components, via the RS  
 CC domains, to form a bridge between the 5' and 3' splice site  
 CC binding components, U1 snRNP and U2AF. Binds to  
 CC putative-rich RNA sequences, either 5'-AGSAGAGTA-3' (S=C or G) or  
 CC 5'-GTTGAGTA-3'. Can bind to beta-globin mRNA and commit it to the  
 CC splicing pathway.  
 CC -1- SUBUNIT: In vitro, binds SFRS1 (ASF/SF2), SNRP70 and U2AF1 but not  
 CC U2AF2. Binds SFRS12.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- PTM: Extensively phosphorylated on serine residues in the RS  
 CC domain.  
 CC -1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M90104; AAA60306.1; -;  
 DR EMBL; X62447; CAA44307.1; -;  
 DR EMBL; X75755; CAA53383.1; -;  
 DR EMBL; BC000339; AAH00339.1; -;  
 DR EMBL; L03693; AAA60162.1; -;  
 DR HSSP; P19339; ISXL;  
 DR Genew; HGNC:10783; SFRS2.  
 DR GK; Q01130; -;  
 DR MIM; 600813; -;  
 GO; GO:0008248; F:pre-mRNA splicing factor activity; TAS.

DR GO:0006397; P:mRNA processing; TAS.  
 DR GO:0006371; P:mRNA splicing; TAS.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rim; 1.  
 DR SMART; SM00360; RRM; 1.  
 DR PROSITE; PS50102; RRM; 1.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 KW Nuclear protein; RNA-binding; mRNA splicing; Phosphorylation.  
 FT DOMAIN 14 92  
 FT DOMAIN 111 116  
 FT DOMAIN 117 221  
 FT CONFLICT 38 38  
 SQ SEQUENCE 221 AA; 25575 MW; 9D1B76BDB5701F5 CRC64;  
 Query Match 3.7%; Score 8; DB 1; Length 221;  
 Best Local Similarity 100.0%; Pred. No. 0.88;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 126 RRRSRSS 133  
 Db 125 RRRSRSS 132  
 RESULT 14  
 SFR2\_MOUSE  
 ID SFR2\_MOUSE STANDARD; PRT; 221 AA.  
 AC Q62093;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Splicing factor, arginine/serine-rich 2 (Splicing factor SC35) (SC-35)  
 DE (Splicing component, 35 kDa) (PR264 protein).  
 GN SFRS2 OR PR264.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98447613; PubMed=9774382;  
 RA Yang L., Embree L.J., Tsai S., Hickstein D.D.;  
 RT "Oncoprotein TLS interacts with serine-arginine proteins involved in  
 RT RNA splicing.";  
 RL J. Biol. Chem. 273:27761-27764(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Mammary gland;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stadelton M., Soares M.B., Bonaldo W.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullighy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.L., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey U., Helton E., Kettelman M., Maman A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Medan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE OF 1-121 FROM N.A.  
 RC STRAIN=129/Sv; TISSUE=Liver;  
 RA Galliard C., Perbal B.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: Necessary for the splicing of pre-mRNA. It is required  
CC for formation of the earliest ATP-dependent splicing complex and  
CC interacts with spliceosomal components bound to both the 5' and  
CC 3' splice sites during spliceosome assembly. It also is required  
CC for ATP-dependent interactions of both U1 and U2 snRNPs with pre-  
CC mRNA (By similarity).  
CC -!- SUBUNIT: In vitro, binds SFRS1 (ASF/SF2), SNRP70 and U2AF1 but not  
CC U2AF2. Binds SFRS12 (By similarity).  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- PTM: Extensively phosphorylated on serine residues in the RS  
CC domain (By similarity).  
CC -!- SIMILARITY: Belongs to the splicing factor SR family.  
CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF077858; AAC71000.1; -  
DR EMBL; BC005493; AAH05493.1; -  
DR EMBL; X98511; CA67134.1; -  
DR HSSP; P19339; 1SXL.  
DR WGD; MG138284; Sfrs2.  
DR GO; GO:0005681; C:spliceosome complex; IDA.  
DR InterPro; IPR000504; C:spliceosome complex; RNA\_rec\_mot.  
DR Pfam; PF00076; rrm; 1.  
DR SMART; SM00360; RRM; 1.  
DR PROSITE; PSS0102; RRM; 1.  
DR PROSITE; PSS0030; RRM\_RNP\_1; 1.  
DR Nuclear protein; RNA-binding; mRNA splicing; Phosphorylation.  
FT DOMAIN 14 92 RNA-BINDING (RRM).  
FT DOMAIN 111 116 GLY-RICH (HINCH REGION).  
FT DOMAIN 117 221 ARG/SER-RICH (RS DOMAIN).  
SQ SEQUENCE 221 AA; 25476 MW; 68121ACAD35714FA CRC64;

Query Match 3.7%; Score 8; DB 1; Length 221;  
Best Local Similarity 100.0%; Pred. No. 0.88;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 RRRSRSS 133  
Db 125 RRRSRSS 132

RESULT 15  
SFR7\_HUMAN  
ID SFR7\_HUMAN STANDARD; PRT; 238 AA.  
AC Q16629;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Splicing factor, arginine/serine-rich 7 (Splicing factor 968).  
GN SFRS7.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 7-24; 30-58 AND 79-87.  
RC TISSUE=Placenta;  
RC MEDLINE=94283389; PubMed=8013463;  
RA Cavaloc Y., Popielarz M., Fuchs J.-P., Gatroni R., Stevenin J.;  
RT "Characterization and cloning of the human splicing factor 968: a  
RT novel 35 kDa factor of the serine/arginine protein family.";  
RL EMBJ. J. 13:2639-2649(1994).  
RU [2]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RC TISSUE=Placenta;  
RC MEDLINE=95355374; PubMed=7629084;

RA Popielarz M., Cavaloc Y., Matei M.-G., Gatroni R., Stevenin J.;  
RT "The gene encoding human splicing factor 968. Structure, chromosomal  
RT location, and expression of alternatively processed transcripts";  
RL J. Biol. Chem. 270:17830-17835(1995).  
RX [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain, Cervix, Prostate, and Skin;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marzina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stjepanovich M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Uscin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.U., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Gilwood J., Schmutz J., Myers R.M.,  
RA Buterfield Y.S.N., Krzywinski M.I., Skalska V., Smallus D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16699-16903(2002).  
RX [4]  
RP INTERACTION WITH CCN1, CDC2L1, AND CDC2L2.  
RX MEDLINE=2249644; PubMed=12501247;  
RA Hu D., Mayeda A., Trembley J.H., Lahti J.M., Kidd V.J.;  
RT "CDK1 complexes promote pre-mRNA splicing";  
RL J. Biol. Chem. 278:8623-8629(2003).  
CC -!- FUNCTION: Required for pre-mRNA splicing. Can also modulate  
CC alternative splicing in vitro.  
CC -!- SUBUNIT: Found in large molecular weight complexes containing  
CC CCN1 and the p110 isoforms of either CDC2L1 or CDC2L2.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Comment=Isoforms, often lacking the RS domain and differentially  
CC expressed in fetal tissues, may be involved in modulation of 968  
CC function;  
CC Name=1;  
CC IsoId=Q16629-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q16629-2; Sequence=VSP\_005872, VSP\_005873;  
CC Name=3;  
CC IsoId=Q16629-3; Sequence=VSP\_005874, VSP\_005875;  
CC -!- TISSUE SPECIFICITY: Brain, liver, kidney and lung.  
CC -!- PTM: Extensively phosphorylated on serine residues in the RS  
CC domain.  
CC -!- SIMILARITY: Belongs to the splicing factor SR family.  
CC -!- SIMILARITY: Contains 1 CCHC-type zinc finger.  
CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L22253; AAA35495.1; -  
DR EMBL; L41887; AAA88098.1; -  
DR EMBL; BC000997; AAH00997.1; -  
DR EMBL; BC017369; AAH17369.1; -  
DR EMBL; BC017908; AAH17908.1; -  
DR EMBL; BC022328; AAH22328.1; -  
DR PIR; AS7198; AS7198.  
DR HSSP; P11940; 1CVT.

```

DR   Genew: HGNC:10789; SFRS7.
DR   GK; Q16629; -.
DR   MIM; 600572; -.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0008248; F:pre-mRNA splicing factor activity; TAS.
DR   GO; GO:0006397; P:mRNA processing; TAS.
DR   GO; GO:0006371; P:mRNA splicing; TAS.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   InterPro; IPR01878; Znf_CCHC.
DR   Pfam; PF00076; xrm; 1.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   SMART; SM00360; RRM; 1.
DR   SMART; SM00343; Znf_C2HC; 1.
DR   PROSITE; PS50102; RRM; 1. 1. FALSE_NEG.
DR   PROSITE; PS00030; RRM_RNP_1;
DR   PROSITE; PS50158; zf_CCHC; 1.
DR   Nuclear protein; RNA-binding; mRNA splicing; Alternative splicing;
DR   Phosphorylation; Repeat; Zinc-finger.
DR   DOMAIN 11 84 RNA-BINDING (RRM).
FT   ZN_FING 104 120 CCHC-TYPE.
FT   DOMAIN 121 238 ARG/SER-RICH (RS DOMAIN).
FT   DOMAIN 153 227 6 X 8 AA REPEATS OF R-R-S-R-S-X-S-X.
FT   REPEAT 153 160 1.
FT   REPEAT 161 168 2.
FT   REPEAT 169 176 3.
FT   REPEAT 177 184 4.
FT   REPEAT 211 218 5 (APPROXIMATE).
FT   REPEAT 219 226 6 (APPROXIMATE).
FT   REPEAT 238 246 SRSRSH -> AENLR (in isoform 2).
FT   VARSPLIC 130 135 /FTId=VSP_005872.
FT   VARSPLIC 136 238 Missing (in isoform 2).
FT   VARSPLIC 130 132 /FTId=VSP_005873.
FT   VARSPLIC 133 238 SRS -> YLF (in isoform 3).
FT   VARSPLIC 133 238 Missing (in isoform 3).
FT   VARSPLIC 133 238 /FTId=VSP_005875.
SQ   SEQUENCE 238 AA; 27366 MW; 49136754D9630853 CRC64;

```

Query Match 3.7%; Score 8; DB 1; Length 238;  
 Best Local Similarity 100.0%; Pred. No. 0.94;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRRS 133  
 DB 125 RRRSRRS 132

Search completed: July 22, 2004, 17:19:21  
 Job time : 22 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 17:13:52 ; Search time 39 Seconds  
(without alignments)  
1747.485 Million cell updates/sec

Title: us-09-787-491b-1

Perfect score: 216  
Sequence: 1 MSRYLRPNTSLFVRNVADD.....PKGKMKVLQYECTNILLTV 216

Scoring table:  
OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL 25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	155	71.8	173	4 Q96G09	Q96G09 homo sapien
2	155	71.8	183	4 Q60572	Q60572 homo sapien
3	155	71.8	183	11 Q70307	Q70307 mus musculi
4	155	71.8	262	4 Q75494	Q75494 homo sapien
5	155	71.8	262	11 Q88468	Q88468 mus musculi
6	153	70.8	183	11 Q9R0T9	Q9R0T9 mus musculi
7	153	70.8	262	11 Q9R0U0	Q9R0U0 mus musculi
8	146	67.6	261	4 Q96P17	Q96P17 homo sapien
9	124	57.4	261	11 Q8CF21	Q8CF21 mus musculi
10	124	57.4	69	11 Q8CF51	Q8CF51 mus musculi
11	27	12.5	261	4 Q8WAF0	Q8WAF0 mus sapien
12	25	11.6	258	13 Q7ZUJ8	Q7ZUJ8 xenopus lae
13	21	9.7	261	4 Q8WV25	Q8WV25 homo sapien
14	19	8.8	248	13 Q7S2S0	Q7S2S0 brachydanto
15	12	5.6	773	5 Q9UJ79	Q9UJ79 drosophila
16	12	5.6	767	5 Q9VEJ1	Q9VEJ1 drosophila

17	9	4.2	226	10 Q8LQJ8	Q8LQJ8 oryza sativ
18	9	4.2	236	4 Q96JW1	Q96JW1 homo sapien
19	9	4.2	283	4 Q8TF00	Q8TF00 homo sapien
20	9	4.2	299	4 Q9Y4N3	Q9Y4N3 homo sapien
21	9	4.2	312	4 Q8NAV1	Q8NAV1 homo sapien
22	9	4.2	312	6 Q8HXH6	Q8HXH6 macaca fasc
23	9	4.2	312	11 Q8BYV2	Q8BYV2 mus musculi
24	9	4.2	347	10 Q9SU74	Q9SU74 arabidopsis
25	9	4.2	521	5 Q9VDR5	Q9VDR5 drosophila
26	9	4.2	523	13 Q803H1	Q803H1 brachydanto
27	9	4.2	805	4 Q8TF01	Q8TF01 homo sapien
28	9	4.2	872	5 Q8M2H3	Q8M2H3 drosophila
29	9	4.2	872	5 Q9VGU5	Q9VGU5 drosophila
30	8	3.7	48	5 Q86LB9	Q86LB9 entamoeba h
31	8	3.7	52	12 Q91G66	Q91G66 epihyas po
32	8	3.7	54	12 Q8UR11	Q8UR11 phthorlinaea
33	8	3.7	55	12 Q7F9U3	Q7F9U3 adoxophyes
34	8	3.7	56	12 Q9DVM6	Q9DVM6 pluteia xy
35	8	3.7	60	12 Q9PYU9	Q9PYU9 xestia c-ni
36	8	3.7	79	12 Q7T5F8	Q7T5F8 anthracia p
37	8	3.7	87	12 Q91EX8	Q91EX8 cydia pomon
38	8	3.7	118	5 Q8MSW0	Q8MSW0 caenorhabdi
39	8	3.7	129	11 Q9CVD6	Q9CVD6 mus musculi
40	8	3.7	141	6 Q8WMQ1	Q8WMQ1 macaca mula
41	8	3.7	146	6 Q8WMQ0	Q8WMQ0 macaca mula
42	8	3.7	153	5 Q8WSW1	Q8WSW1 caenorhabdi
43	8	3.7	155	4 Q8WV22	Q8WV22 homo sapien
44	8	3.7	155	11 Q8K194	Q8K194 mus musculi
45	8	3.7	160	11 Q9CW27	Q9CW27 mus musculi

## ALIGNMENTS

## RESULT 1

Q96G09 ID Q96G09 PRELIMINARY; PRT; 173 AA.  
AC Q96G09;  
DT 01-DEC-2001 (TREMURel. 19, Created)  
DT 01-DEC-2001 (TREMURel. 19, Last sequence update)  
DT 01-OCT-2003 (TREMURel. 25, Last annotation update)  
DE T1S-associated serine-arginine protein 1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
CX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Uterus;  
RA Strauberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC010074; AAI10074.1; ~  
DR GO: GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro: IPR000504; RNA\_rec\_mot.  
DR Pfam: PF00076; trm; 1.  
DR SMART: SM00360; RRM; 1.  
DR PROSITE: PSS0102; RRM; 1.  
DR PROSITE: PSS0030; RRM\_RNP\_1; 1.  
SQ SEQUENCE 173 AA; 21000 MW; 4AB40FBE24846495 CRC64;

Query Match 71.8%; Score 155; DB 4; Length 173;  
Best Local Similarity 100.0%; Pred. No. 1.3e-158;  
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSRYLRPNTSLFVRNVADTRSEDI	RRERGRYGPYDVYVPI	LDFTRRRRGPAYVQFED	60
DB	1	MSRYLRPNTSLFVRNVADTRSEDI	RRERGRYGPYDVYVPI	LDFTRRRRGPAYVQFED	60
QY	61	VRAEDALHLIDKRWICGRQIEITQ	AGGDKTGNOKAKGGRVYSSR	VDYDRYSR	120
DB	61	VRAEDALHLIDKRWICGRQIEITQ	AGGDKTGNOKAKGGRVYSSR	VDYDRYSR	120
QY	121	SRSYERRRSRSRSPDYNRYSYS	PRNSRPTGRDR	155	



Db 121 SRSYERRSRSPDYNNRYSYSPNSRPTGRPR 155

RESULT 2

060572 PRELIMINARY; PRT; 183 AA.

AC 060572; TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-AUG-1998 (TREMBLrel. 24, Last annotation update)

DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)

DE TLS-associated serine-arginine protein TASR (Hypothetical protein FLJ10424)

DE (TAS-associated serine-arginine protein 1) (TLS-associated protein

DE TASR-1) (TLS-associated SR protein 1) (Splicing factor SRp38-2).

GN TASR.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

LN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Leukemia;

RA Yang L., Embree L., Tsai S., Hickestein D.D.;

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

LN [2]

RP SEQUENCE FROM N.A.

RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,

RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,

RA Wagatsuma M., Hoshida T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,

RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,

RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,

RA Yamamoto J., Makamatsu Y., Nakamura Y., Nagahara K., Masuko Y.,

RA Ninomiya K., Iwayanagi T.;

RT "NEDO human cDNA sequencing project.;"

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

LN [3]

RP SEQUENCE FROM N.A.

RC TISSUE=Placenta;

RA Strausberg R.;

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

LN [4]

RP SEQUENCE FROM N.A.

RA Clinton J.M., Chansky H.A., Odell D.D., Zielinska-Kwiatkowska A.,

RA Yang L.;

RT "Characterization and expression of the human gene encoding two TLS-

RT associated serine-arginine (TASR) proteins.;"

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

LN [5]

RP SEQUENCE FROM N.A.

RA MEDLINE=98447613; PubMed=9774382;

RA Yang L., Embree L.J., Tsai S., Hickestein D.D.;

RT "Oncoprotein TLS interacts with serine-arginine proteins involved in

RT RNA splicing.;"

RL J. Biol. Chem. 273:27761-27764(1998).

LN [6]

RP SEQUENCE FROM N.A.

RA MEDLINE=20242028; PubMed=10779324;

RA Yang L., Embree L.J., Hickestein D.D.;

RT "TUS-ERG leukemia fusion protein inhibits RNA splicing mediated by

RT serine-arginine proteins.;"

RL Mol. Cell. Biol. 20:3345-3354(2000).

LN [7]

RP SEQUENCE FROM N.A.

RA Clinton J.M., Chansky H.A., Odell D.D., Zielinska-Kwiatkowska A.,

RA Yang L.;

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

LN [8]

RP SEQUENCE FROM N.A.

RA MEDLINE=2206915; PubMed=12419250;

RA Shin C., Manley J.L.;

RT "The SR Protein SRP38 Represses Splicing in M Phase Cells.;"

RL Cell 111:407-417(2002).

DR EMBL; AF047448; AAC70918.1; -.

DR EMBL; AK001286; BA591601.1; -

DR EMBL; BC001107; AA01107.1; -

DR EMBL; AF419331; AAL16665.1; -

DR EMBL; AY048592; AAL106098.1; -

DR EMBL; AY150181; AAN65381.1; -

DR HSSP; P11940; ICVJ.

DR GO; GO:0008248; F:pre-mRNA splicing factor activity; TAS.

DR GO; GO:0006397; P:mRNA processing; TAS.

DR GO; GO:0000245; P:spliceosome assembly; TAS.

DR InterPro; IPR000504; RNA\_rec\_mot.

DR Pfam; PF00076; rtm; 1.

DR SMART; SM00360; RRM; 1.

DR PROSITE; PS0102; RRM; 1.

DR PROSITE; PS00030; RRM\_RNP\_1; 1.

DR Hypothetical protein.

SW SEQUENCE 183 AA; 22222 MW; 4AA87CA9B51A131 CRC64;

Query Match 71.8%; Score 155; DB 4; Length 183;

Best Local Similarity 100.0%; Pred. No. 1.4e-158;

Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNNVADDTRESDELRRFGRYGPIVDVYPLDFTYRPRGFAVYQFED 60

DB 1 MSRYLRPNTSLFVNNVADDTRESDELRRFGRYGPIVDVYPLDFTYRPRGFAVYQFED 60

QY 61 VRDAEDALHNLDRKWCIGROIEIQPAGDRKTPNQMKAKGRNYSRSDYDRYRSR 120

DB 61 VRDAEDALHNLDRKWCIGROIEIQPAGDRKTPNQMKAKGRNYSRSDYDRYRSR 120

QY 121 SRSYERRSRSPDYNNRYSYSPNSRPTGRPR 155

DB 121 SRSYERRSRSPDYNNRYSYSPNSRPTGRPR 155

RESULT 3

070307 PRELIMINARY; PRT; 183 AA.

AC 070307;

DT 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE TLS-associated protein with SR repeats.

GN NSSR.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;

LN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=98447613; PubMed=9774382;

RA Yang L., Embree L.J., Tsai S., Hickestein D.D.;

RT "Oncoprotein TLS interacts with serine-arginine proteins involved in

RT RNA splicing.;"

RL J. Biol. Chem. 273:27761-27764(1998).

DR EMBL; AF042383; AAC70916.1; -.

DR HSSP; P11940; ICVJ.

DR MGD; MGI:1333805; Nestr.

DR GO; GO:0003676; F:nucleic acid binding; IEA.

DR InterPro; IPR000504; RNA\_rec\_mot.

DR Pfam; PF00076; rtm; 1.

DR PROSITE; PS0102; RRM; 1.

DR PROSITE; PS00030; RRM\_RNP\_1; 1.

SW SEQUENCE 183 AA; 22222 MW; 4AA87CA9B51A131 CRC64;

Query Match 71.8%; Score 155; DB 11; Length 183;

Best Local Similarity 100.0%; Pred. No. 1.4e-158;

Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNNVADDTRESDELRRFGRYGPIVDVYPLDFTYRPRGFAVYQFED 60

DB 1 MSRYLRPNTSLFVNNVADDTRESDELRRFGRYGPIVDVYPLDFTYRPRGFAVYQFED 60

QY 61 VRDAEDALHNLDRKWCIGROIEIQPAGDRKTPNQMKAKGRNYSRSDYDRYRSR 120

```

Db      61 VRDAEDALHNLDRKWMICGRQIEIQFAGDRTPTQMKAKGGRNYSSSRDYDYDRYRSR 120
QY      121 SRSYERRRSRSPDYVNRYSRSPRNSRPTGRPR 155
      121 SRSYERRRSRSPDYVNRYSRSPRNSRPTGRPR 155
Db
RESULT 4
075494 PRELIMINARY; PRT; 262 AA.
ID 075494
AC 075494;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE TLS-associated protein TASR-2 (TLS-associated serine-arginine protein
  2) (TLS-associated SR protein 2) (SRp40) (Splicing factor
  SRp38).
DE TASR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Leukemia;
RA Yang L., Embree L., Hickstein D.D., Tsai S.;
RT "Molecular cloning of human TASR-2, a TLS-associated protein with Ser-
  Arg repeats.";
RT Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.
RC MEDLINE=98447613; PubMed=9774382;
RA Yang L., Embree L.J., Tsai S., Hickstein D.D.;
RT "Oncoprotein TLS interacts with serine-arginine proteins involved in
  RNA splicing.";
RT J. Biol. Chem. 273:27761-27764(1998).
RN 14
RP SEQUENCE FROM N.A.
RC MEDLINE=20242028; PubMed=10779324;
RA Yang L., Embree L.J., Hickstein D.D.;
RT "TLS-ERG leukemia fusion protein inhibits RNA splicing mediated by
  serine-arginine proteins.";
RT Mol. Cell. Biol. 20:3345-3354(2000).
RN 15
RP SEQUENCE FROM N.A.
RA Clinton J.M., Chansky H.A., Odell D.D., Zielinska-Kwiatkowska A.,
  Yang L.;
RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN 16
RP SEQUENCE FROM N.A.
RC MEDLINE=2163667; PubMed=11684676;
RA Comper A.E., Caceres J.F., Mayeda A., Screaton G.R.;
RT "Serine-Arginine (SR) Protein-like Factors That Antagonize Authentic
  SR Proteins and Regulate Alternative Splicing.";
RT J. Biol. Chem. 276:48908-48914(2001).
RN 17
RP SEQUENCE FROM N.A.
RC MEDLINE=22306915; PubMed=12419250;
RA Shin C., Manley J.L.;
RT "The SR Protein SRP38 Represses Splicing in M Phase Cells.";
RT Cell 111:407-417(2002).
DR EMBL: AF067730; AAC26727.1; -
DR EMBL: BC005039; AAH05039.1; -
DR EMBL: AY048592; AAL06099.1; -
DR EMBL: AF449427; AAL57514.1; -
DR EMBL: AY150180; AAN65380.1; -
DR HSSP: P11940; ILCVT.

```

```

DR Genew; HGNC:16713; FUSIP1.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 262 AA; 31300 MW; 205F9D36CBFB4 CRC64;
Query Match 71.8%; Score 155; DB 4; Length 262;
Best Local Similarity 100.0%; Pred. No. 1.9e-158;
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSRYLRPNTSLFVRNADDTSRSDLRERFGXGPIVDVYVPLDFYTRRRPGFAYVQED 60
      1 MSRYLRPNTSLFVRNADDTSRSDLRERFGXGPIVDVYVPLDFYTRRRPGFAYVQED 60
Db 61 VRDAEDALHNLDRKWMICGRQIEIQFAGDRTPTQMKAKGGRNYSSSRDYDYDRYRSR 120
QY 61 VRDAEDALHNLDRKWMICGRQIEIQFAGDRTPTQMKAKGGRNYSSSRDYDYDRYRSR 120
Db 61 VRDAEDALHNLDRKWMICGRQIEIQFAGDRTPTQMKAKGGRNYSSSRDYDYDRYRSR 120
QY 121 SRSYERRRSRSPDYVNRYSRSPRNSRPTGRPR 155
      121 SRSYERRRSRSPDYVNRYSRSPRNSRPTGRPR 155
Db 121 SRSYERRRSRSPDYVNRYSRSPRNSRPTGRPR 155
RESULT 5
088468 PRELIMINARY; PRT; 262 AA.
ID 088468
AC 088468;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE TLS-associated protein TASR-2 (FUS-associated protein with
  serine-arginine repeats) (Similar to neural-salient
  serine/arginine-rich).
DE NSR OR FASR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RA Yang L., Embree L., Tsai S., Hickstein D.D.;
RT "Molecular cloning of TASR-2, a TLS-associated protein with Ser-Arg
  repeats.";
RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217951;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
  Arakawa T., Hara A., Fukunishi Y., Komno H., Adachi J., Fukuda S.,
  Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
  Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
  Kadota K., Matsuda Y., Ashburner M., Batalov S., Casavant T.,
  Pleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
  Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
  Schriml L.M., Straubli F., Suzuki R., Tomita M., Wagner J., Mashio T.,
  Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
  Blake J., Boiffelli D., Bojunga N., Carlinici P., de Bonaldo M.F.,
  Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
  Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
  Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
  Nordone P., Ringwald M., Rodriguez I., Sakamoto N.,
  Sasaki H., Sato K., Schoenbach C., Seya T., Shibata R., Storch K.-F.,
  Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
  Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohatsu S.,
  Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
RN 13
RP SEQUENCE FROM N.A.

```

```
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF060490; AAC26715.1; -.
DR EMBL; AK014345; BAB29286.1; -.
DR EMBL; BC043060; AAA43060.1; -.
DR HSSP; P11940; ICVJ.
DR MGD; MGI:1333805; Nsr.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR PROSITE; PS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 262 AA; 3130 MW; 205F95D36CBFB4 CRC64;

Query Match 71.8%; Score 155; DB 11; Length 262;
Best Local Similarity 100.0%; Pred. No. 1.9e-158;
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNNVADPTRSEDLRRFGRYGPIVVVYVPLDYTRPRGFAVVOED 60
DB 1 MSRYLRPNTSLFVNNVADPTRSEDLRRFGRYGPIVVVYVPLDYTRPRGFAVVOED 60

QY 61 VRDAEDALHNDLRKWCGRQIEIQPAGCDRKTNPOMKAKEGRNVYSSSRYYDDYDRYRSR 120
DB 61 VRDAEDALHNDLRKWCGRQIEIQPAGCDRKTNPOMKAKEGRNVYSSSRYYDDYDRYRSR 120

QY 121 SRSYERRRSRSGSPDYNNRSYSPNRSRPTGR 155
DB 121 SRSYERRRSRSGSPDYNNRSYSPNRSRPTGR 155

RESULT 6
Q9ROT9 PRELIMINARY; PRT; 183 AA.
ID Q9ROT9;
AC Q9ROT9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Neutral specific sr protein NSSR 2.
DE NSSR OR NSSR 2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_Taxid=10090;
CX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=20051033; PubMed=10583508;
RA Komatsu M., Komimami E., Arahata K., Tsukahara T.;
RT "Cloning and characterization of two neutral-salient serine/arginine-
RT rich (NSSR) proteins involved in the regulation of alternative
RT splicing in neurons.";
RL Genes Cells 4:593-608 (1999).
DR EMBL; AB015895; BAA35093.1; -.
DR HSSP; P11940; ICVJ.
DR MGD; MGI:1333805; Nsr.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 183 AA; 22167 MW; 4AA87CA9B57C157 CRC64;

Query Match 70.8%; Score 153; DB 11; Length 183;
Best Local Similarity 100.0%; Pred. No. 2e-156;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNNVADPTRSEDLRRFGRYGPIVVVYVPLDYTRPRGFAVVOED 60
DB 1 MSRYLRPNTSLFVNNVADPTRSEDLRRFGRYGPIVVVYVPLDYTRPRGFAVVOED 60

QY 61 VRDAEDALHNDLRKWCGRQIEIQPAGCDRKTNPOMKAKEGRNVYSSSRYYDDYDRYRSR 120
DB 61 VRDAEDALHNDLRKWCGRQIEIQPAGCDRKTNPOMKAKEGRNVYSSSRYYDDYDRYRSR 120

QY 121 SRSYERRRSRSGSPDYNNRSYSPNRSRPTGR 155
DB 121 SRSYERRRSRSGSPDYNNRSYSPNRSRPTGR 155
```

```
DB 61 VRDAEDALHNDLRKWCGRQIEIQPAGCDRKTNPOMKAKEGRNVYSSSRYYDDYDRYRSR 120
QY 121 SRSYERRRSRSGSPDYNNRSYSPNRSRPTGR 153
DB 121 SRSYERRRSRSGSPDYNNRSYSPNRSRPTGR 153

RESULT 7
Q9ROT0 PRELIMINARY; PRT; 262 AA.
ID Q9ROT0;
AC Q9ROT0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Neutral specific sr protein NSSR 1.
DE NSSR OR NSSR 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_Taxid=10090;
CX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=20051033; PubMed=10583508;
RA Komatsu M., Komimami E., Arahata K., Tsukahara T.;
RT "Cloning and characterization of two neutral-salient serine/arginine-
RT rich (NSSR) proteins involved in the regulation of alternative
RT splicing in neurons.";
RL Genes Cells 4:593-608 (1999).
DR EMBL; AB015894; BAA35092.1; -.
DR HSSP; P11940; ICVJ.
DR MGD; MGI:1333805; Nsr.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 262 AA; 31245 MW; EB32279A4B5B8353 CRC64;

Query Match 70.8%; Score 153; DB 11; Length 262;
Best Local Similarity 100.0%; Pred. No. 2.8e-156;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNNVADPTRSEDLRRFGRYGPIVVVYVPLDYTRPRGFAVVOED 60
DB 1 MSRYLRPNTSLFVNNVADPTRSEDLRRFGRYGPIVVVYVPLDYTRPRGFAVVOED 60

QY 61 VRDAEDALHNDLRKWCGRQIEIQPAGCDRKTNPOMKAKEGRNVYSSSRYYDDYDRYRSR 120
DB 61 VRDAEDALHNDLRKWCGRQIEIQPAGCDRKTNPOMKAKEGRNVYSSSRYYDDYDRYRSR 120

QY 121 SRSYERRRSRSGSPDYNNRSYSPNRSRPTGR 153
DB 121 SRSYERRRSRSGSPDYNNRSYSPNRSRPTGR 153

RESULT 8
Q96P17 PRELIMINARY; PRT; 261 AA.
ID Q96P17;
AC Q96P17;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE TIS-associated protein TISR-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_Taxid=9606;
CX [1]
RN SEQUENCE FROM N.A.
RA Clinton J.M., Chansky H.A., Odell D.D., Zielinska-Kwiatkowska A.,
Yang L.;
```

RT "Characterization and expression of the human gene encoding two TLS-  
RT associated serine-arginine (TSAR) proteins."  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF413332; AAL1666.1; "  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF00076; rrm; 1.  
DR SMART; SM00360; RRM; 1.  
DR PROSITE; PS50102; RRM; 1.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
SQ SEQUENCE 261 AA; 31213 MW; A07499B1D4C7570 CRC64;

Query Match 67.6%; Score 146; DB 4; Length 261;  
Best Local Similarity 100.0%; Pred. No. 9,7e-149;  
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNADVADTRSEDLRRFGYGVIVVYVLDYTRRPGFAYVQED 60  
DB 1 MSRYLRPNTSLFVNADVADTRSEDLRRFGYGVIVVYVLDYTRRPGFAYVQED 60  
QY 61 VRDAEDALHNLDKMTICGROIIEQAGCDKRTPNQMKAKGRNVYSSSRDYDYRRSR 120  
DB 61 VRDAEDALHNLDKMTICGROIIEQAGCDKRTPNQMKAKGRNVYSSSRDYDYRRSR 120  
QY 121 SRSYERRSRSRSPDYNYRRSGYSPRN 146  
DB 121 SRSYERRSRSRSPDYNYRRSGYSPRN 146

RESULT 9  
Q8CFZ1 PRELIMINARY; PRT; 261 AA.  
ID Q8CFZ1  
AC Q8CFZ1  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Similar to FUS interacting protein (Serine-arginine rich) 1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVB/N-3;  
RA Strausberg R.;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC037591; AAH37591.1; "  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF00076; rrm; 1.  
DR SMART; SM00360; RRM; 1.  
DR PROSITE; PS50102; RRM; 1.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
SQ SEQUENCE 261 AA; 31185 MW; C9500BF970FAAE1D CRC64;

Query Match 57.4%; Score 124; DB 11; Length 261;  
Best Local Similarity 100.0%; Pred. No. 5,2e-125;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 SEDLRREFGRRYGVIVVYVLDYTRRPGFAYVQEDVADALHNLDKMTICGROI 82  
DB 23 SEDLRREFGRRYGVIVVYVLDYTRRPGFAYVQEDVADALHNLDKMTICGROI 82  
QY 83 IQPAQGDRTPNQMKAKGRNVYSSSRDYDYRRSRSGYERRSRSGYFYNYRSY 142  
DB 83 IQPAQGDRTPNQMKAKGRNVYSSSRDYDYRRSRSGYERRSRSGYFYNYRSY 142  
QY 143 SPBN 146  
DB 143 SPBN 146

RESULT 10

Q8CF51 PRELIMINARY; PRT; 69 AA.  
ID Q8CF51  
AC Q8CF51  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Neural-salient serine/arginine-rich.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573 (2002).  
DR EMBL; AK005295; BAC25110.1; "  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF00076; rrm; 1.  
DR SMART; SM00360; RRM; 1.  
DR PROSITE; PS50102; RRM; 1.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
SQ SEQUENCE 69 AA; 8347 MW; AD2322054AE747B5 CRC64;

Query Match 26.4%; Score 57; DB 11; Length 69;  
Best Local Similarity 100.0%; Pred. No. 2,8e-53;  
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNADVADTRSEDLRRFGYGVIVVYVLDYTRRPGFAYVQ 57  
DB 1 MSRYLRPNTSLFVNADVADTRSEDLRRFGYGVIVVYVLDYTRRPGFAYVQ 57

RESULT 11  
Q8WYF0 PRELIMINARY; PRT; 261 AA.  
ID Q8WYF0  
AC Q8WYF0  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Strp3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21623667; PubMed=11684676;  
RA Cowper A.E., Caceres J.F., Mayeda A., Screaton G.R.;  
RT "Serine-Arginine (SR) Protein-like Factors That Antagonize Authentic  
RT SR Proteins and Regulate Alternative Splicing."  
RL J. Biol. Chem. 276:48908-48914 (2001).  
DR EMBL; AF449428; AAL57515.1; "  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF00076; rrm; 1.  
DR SMART; SM00360; RRM; 1.  
DR PROSITE; PS50102; RRM; 1.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
SQ SEQUENCE 261 AA; 30512 MW; C1473A5760462B54 CRC64;

Query Match 12.5%; Score 27; DB 4; Length 261;  
Best Local Similarity 100.0%; Pred. No. 2,1e-20;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 PLDFYTRRRPGFAYVQEDVRDAEDAL 68  
DB 42 PLDFYTRRRPGFAYVQEDVRDAEDAL 68

## RESULT 12

07ZTJ8 PRELIMINARY; PRT; 258 AA.  
 AC 07ZTJ8  
 DT 01-JUN-2003 (TRENBLrel. 24, Created)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Similar to FUS interacting protein (Serine-arginine rich) 1.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;  
 CC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8335;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA Klein S., Strausberg R.;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC046695; AAH46695.1;  
 DR GO; GO:0003676; F:mucleic acid binding; IEA.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 1.  
 DR SMART; SM00360; RRM; 1.  
 DR PROSITE; PS50102; RRM; 1.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 DR SEQUENCE 258 AA; 30285 MW; 8436A11BE3630BAB CRC64;

Query Match 11.6%; Score 25; DB 13; Length 258;  
 Best Local Similarity 100.0%; Pred. No. 3e-18;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 RPRGFAYVQFEDVDAEDALHND 72  
 DB 48 RPRGFAYVQFEDVDAEDALHND 72

## RESULT 13

08MW25 PRELIMINARY; PRT; 261 AA.  
 AC 08MW25  
 DT 01-MAR-2002 (TRENBLrel. 20, Created)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)  
 DE Similar to FUS interacting protein (Serine-arginine rich) 2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Strausberg R.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC021715; AAH21715.1;  
 DR GO; GO:0003676; F:mucleic acid binding; IEA.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 1.  
 DR SMART; SM00360; RRM; 1.  
 DR PROSITE; PS50102; RRM; 1.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 DR SEQUENCE 261 AA; 30454 MW; 2420CBDE4284755 CRC64;

Query Match 9.7%; Score 21; DB 4; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-14;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 PLDFYTRRRGFAVQFEDVR 62  
 DB 42 PLDFYTRRRGFAVQFEDVR 62

## RESULT 14

07SZSO PRELIMINARY; PRT; 248 AA.  
 AC 07SZSO  
 DT 01-OCT-2003 (TRENBLrel. 25, Created)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 CC Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diachenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Frange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Boulford G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,  
 RA Krzyzanski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Maira M.A.;  
 RT "Genes and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA Strausberg R.;  
 RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC056275; AAH56275.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 248 AA; 29357 MW; C6E243ACD459335 CRC64;

Query Match 8.8%; Score 19; DB 13; Length 248;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-12;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 QFEDVDAEDALHNDKRW 75  
 DB 57 QFEDVDAEDALHNDKRW 75

## RESULT 15

09UJ9 PRELIMINARY; PRT; 773 AA.  
 AC 09UJ9  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE SFI protein.  
 GN SFI OR CG5836.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Mazroui R., Puoti A., Kramer A.;  
 RT "Splicing factor SFI from Drosophila and Caenorhabditis contains an N-

RT terminal domain not present in the mammalian or yeast homologues."  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 1 KH DOMAIN.  
DR EMBL; AJ243904; CAB64937.1; -.  
DR FlyBase; FBgn0025571; SPL.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR004087; KH\_dom.  
DR InterPro; IPR004088; KH\_type\_1.  
DR InterPro; IPR01878; Znf\_CCHC.  
DR Pfam; PF00013; KH; 1.  
DR Pfam; PF00098; zf-CCHC; 1.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR SMART; SM00322; KH; 1.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS50084; KH\_TYPR\_1; 1.  
DR PROSITE; PS50158; ZF\_CCHC; 1.  
SQ SEQUENCE 773 AA; 86526 MW; BEFCD08AE4755098 CRC64;

Query Match 5.6%; Score 12; DB 5; Length 773;  
Best local similarity 100.0%; Pred.No. 0.00085;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 RRSRSRSYERRR 128  
DB 145 RRSRSRSYERRR 156

Search completed: July 22, 2004, 17:20:03  
Job time : 41 secs

**This Page Blank (uspto)**









```

QY 1021 AATAAAGTTCCATTTTCCATTTATGGCAATCTTGCCTTTGTTATTTTGGTCC 1080
DB 1021 AATAAAGTTCCATTTTCCATTTATGGCAATCTTGCCTTTGTTATTTTGGTCC 1080
QY 1081 AGTGTTCCTGCTTAATCATTTGCTTTGTGGCACTGTGTTTATTTACTGTGACACAC 1140
DB 1081 AGTGTTCCTGCTTAATCATTTGCTTTGTGGCACTGTGTTTATTTACTGTGACACAC 1140
QY 1141 ATGCAGTTTACATCTGTCTTAACCTCTTCCAGTAAATTCATTAATTTGACAT 1200
DB 1141 ATGCAGTTTACATCTGTCTTAACCTCTTCCAGTAAATTCATTAATTTGACAT 1200
QY 1201 CCAGTAAAGAGGCGCCATCTCTCCACCTCTTCCAGTAAATTCATTAATTTGACAT 1260
DB 1201 CCAGTAAAGAGGCGCCATCTCTCCACCTCTTCCAGTAAATTCATTAATTTGACAT 1260
QY 1261 TATTGAGCCCTTACTGCTGGGCAATCATTTGTAAGTAAATTTGAAATTAATTC 1320
DB 1261 TATTGAGCCCTTACTGCTGGGCAATCATTTGTAAGTAAATTTGAAATTAATTC 1320
QY 1321 CCTTATTCAGTAATGTCTACTGAGCAATCTAGTATCATTTACAGTATGCTCAT 1380
DB 1321 CCTTATTCAGTAATGTCTACTGAGCAATCTAGTATCATTTACAGTATGCTCAT 1380
QY 1381 GTTTGTTGAGGTGTGTTATTCATTAATTTTACACCATTCGTCAATGTAATTA 1440
DB 1381 GTTTGTTGAGGTGTGTTATTCATTAATTTTACACCATTCGTCAATGTAATTA 1440
QY 1441 TAGAACAATATATGATCAAGATTAAGTAAATTTGTTGTTATCTGCCATTTAAAGTAT 1500
DB 1441 TAGAACAATATATGATCAAGATTAAGTAAATTTGTTGTTATCTGCCATTTAAAGTAT 1500
QY 1501 CCAATTTTGTATCATTTATTAATTAATGAAAAATGATTTATCTGTATTAACCTGG 1560
DB 1501 CCAATTTTGTATCATTTATTAATTAATGAAAAATGATTTATCTGTATTAACCTGG 1560
QY 1561 TTATTTGTCAGTGTCTGTAATATCTAGAGTAAATTAATTTGTTATCTGCCCTCAC 1620
DB 1561 TTATTTGTCAGTGTCTGTAATATCTAGAGTAAATTAATTTGTTATCTGCCCTCAC 1620
QY 1621 AAACACATGCTAGATATTAACCCCAAAATAGATTTTAACTTTGATTAAGTATTAAG 1680
DB 1621 AAACACATGCTAGATATTAACCCCAAAATAGATTTTAACTTTGATTAAGTATTAAG 1680
QY 1681 AGACTGGGTCTATTAATTAATTTTGAAGCAACAGAGCTGTATCTTAATCTAT 1740
DB 1681 AGACTGGGTCTATTAATTAATTTTGAAGCAACAGAGCTGTATCTTAATCTAT 1740
QY 1741 TTAGTATGTTCTGTAATTAAGAAATGTTCAACAATTAATCTTTTATGATTAACATG 1800
DB 1741 TTAGTATGTTCTGTAATTAAGAAATGTTCAACAATTAATCTTTTATGATTAACATG 1800
QY 1801 TACATTTTAAAGGGACATGTTCTGTGTAATGAGAAATTAATCTTTTATGATTAACATG 1860
DB 1801 TACATTTTAAAGGGACATGTTCTGTGTAATGAGAAATTAATCTTTTATGATTAACATG 1860
QY 1861 AAAAAAAAAA 1872
DB 1861 AAAAAAAAAA 1872

```

RESULT 3

ABL87934 standard, DNA; 1838 BP.

ABL87934;

17-MAY-2002 (first entry)

Human ovarian cancer related DNA clone SEQ ID NO:10912.

Human; ovarian cancer; ovarian tumour; cytostatic; gene; ds.

```

OS Homo sapiens.
XX
XX WO200192581-A2.
XX
XX 06-DEC-2001.
XX
XX 29-MAY-2001; 2001WO-US017756.
XX
XX 26-MAY-2000; 2000US-0207484P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Algate PA, Harlocker SL, Jones R;
XX
XX WPI; 2002-122075/16.
XX
XX
XX Composition for therapy and diagnosis of ovarian cancer comprising
XX polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
XX polypeptide, antibody specific to polypeptide or T cell expressing
XX polypeptide.
XX
XX Claim 1; SEQ ID NO 10912; 489bp; English.
XX
XX
XX The present invention describes a composition (I) comprising: carriers
XX and immunostimulants; and a polypeptide (II) of a ovarian tumour
XX polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
XX from the 10912 nucleotide sequences as given in ABL87934 to ABL87934,
XX (III) encoding (II) having a sequence (S2), a T cell population of (II),
XX or antigen presenting cells that express (II). (I) has cytostatic
XX activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
XX detecting ovarian cancer in a patient's biological sample preferably
XX serum or ovarian tissue. The method comprises contacting a biological
XX sample from a patient with (IV), detecting the amount of polynucleotide
XX hybridising to (IV) and comparing the amount to a predetermined cutoff
XX value and thereby detecting ovarian cancer in the patient, where the
XX amount of polynucleotide hybridising to (IV) is detected preferably by
XX polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
XX useful for stimulating and/or expanding T cells specific for an ovarian
XX tumour protein comprising contacting T cells with (III) or (II). (III) is
XX useful in design and preparation of ribozyme molecules for inhibiting
XX expression of the tumour polypeptides and proteins in tumour cells; and
XX to isolate a full length gene from a suitable library e.g., a tumour cDNA
XX library using well known techniques
XX
XX Sequence 1838 BP; 563 A; 330 C; 386 G; 559 T; 0 U; 0 Other;
XX
XX Query Match 91.7%; Score 1716; DB 6; Length 1838;
XX Best Local Similarity 99.9%; Pred. No. 0;
XX Matches 1836; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
XX
XX 22 GAGGCTGGCGCGCTGAGCGCGGAGCGGTTGCTGAGCCCGTTAGTGCGCGCGGAG 81
XX 1 GAGGCTGGCGCGCTGAGCGCGGAGCGGTTGCTGAGCCCGTTAGTGCGCGCGGAG 60
XX
XX 82 CACGCGCGCGCATATCCCGTACCTGCGTCCCGCAACACGCTCTGTGTCAGAAAC 141
XX 61 CACGCGCGCGCATATCCCGTACCTGCGTCCCGCAACACGCTCTGTGTCAGAAAC 120
XX
XX 142 GTGCGCGGACGACACGAGTGTGAAGACTTGGGCGGTGAATTTGTGCTTATGCTCTATA 201
XX 121 GTGCGCGGACGACACGAGTGTGAAGACTTGGGCGGTGAATTTGTGCTTATGCTCTATA 180
XX
XX 202 GTTATGTTGTAATGTTCCACTGATTTTACACTGCGCGTCCCAAGAGATTTGCTATGTT 261
XX 181 GTTATGTTGTAATGTTCCACTGATTTTACACTGCGCGTCCCAAGAGATTTGCTATGTT 240
XX
XX 262 CAATTTGAGAGTGTTCGTATGCTGATGCTGAAGACGCTTTACATTAATTTGACAGAAATGATTT 321
XX 241 CAATTTGAGAGTGTTCGTATGCTGATGCTGAAGACGCTTTACATTAATTTGACAGAAATGATTT 300
XX
XX 322 TGTGAGCGGCGAGATTTGAATATACGTTTCCCGAGGGGGATTCGAAAGACCAACCAATACGATG 381
XX 301 TGTGAGCGGCGAGATTTGAATATACGTTTCCCGAGGGGGATTCGAAAGACCAACCAATACGATG 360
XX

```

QY 382 AAAGCGAAGAGGAGGAAATGTGTACGTTCTTACGCTATGATGATTAAGACAGATAC 441  
 DB 361 AAAGCGAAGAGGAGGAAATGTGTACGTTCTTACGCTATGATGATTAAGACAGATAC 420  
 QY 442 AGAGCTTCTAGAGCGGAAAGTTATGAAGAGGAGGAGATCAAGAGCGGCTTTTGATTAC 501  
 DB 421 AGAGCTTCTAGAGCGGAAAGTTATGAAGAGGAGGAGATCAAGAGCGGCTTTTGATTAC 480  
 QY 502 AACTATAGAGATCGTATAGTCTAGAAACAGTAGACCGACTGGAAGACGAGGCGTAG- 560  
 DB 481 AACTATAGAGATCGTATAGTCTAGAAACAGTAGACCGACTGGAAGACGAGGCGTAG- 540  
 QY 561 AGAAGCCATTCCGACATGATAGACCAAACTGCACTGGAAATACCAGTACAGTTCTGCT 620  
 DB 541 AGAAGCCATTCCGACATGATAGACCAAACTGCACTGGAAATACCAGTACAGTTCTGCT 600  
 QY 621 TACTACACTTCAAAAGAGATCTGAAAGCGGAAAAAGAACCAAAAGAGGCGAGTTCAAGC 680  
 DB 601 TACTACACTTCAAAAGAGATCTGAAAGCGGAAAAAGAACCAAAAGAGGCGAGTTCAAGC 660  
 QY 681 ACCAAGGGTGGGTGGAAGGTGCTGCAGTATGAATACTGTACGAATATTTTGACTCTGCT 740  
 DB 661 ACCAAGGGTGGGTGGAAGGTGCTGCAGTATGAATACTGTACGAATATTTTGACTCTGCT 720  
 QY 741 CTGAAAAGTAAAGAAATGTTATGAAAACCTACATGAAATATTTGAAGTCCCTTCAAGTT 800  
 DB 721 CTGAAAAGTAAAGAAATGTTATGAAAACCTACATGAAATATTTGAAGTCCCTTCAAGTT 780  
 QY 801 TGAAGTAAAGATTTTGGACAAATTAAGGAATTAACCTTGTACTGTGGAACTAA 860  
 DB 781 TGAAGTAAAGATTTTGGACAAATTAAGGAATTAACCTTGTACTGTGGAACTAA 840  
 QY 861 TCCCTAAATATGAATAGGTTATATGATTATGAGGTACAGGTCCATAATTAATATTG 920  
 DB 841 TCCCTAAATATGAATAGGTTATATGATTATGAGGTACAGGTCCATAATTAATATTG 900  
 QY 921 GAAACTAGATGTCTGAATATCAAGAAAGACGCATAGTCTCTTACAGTGCCTCTGTTG 980  
 DB 901 GAAACTAGATGTCTGAATATCAAGAAAGACGCATAGTCTCTTACAGTGCCTCTGTTG 960  
 QY 981 GTCTGTCTAAACGAAATGGGTGGGAAAAAGTATGGTCCAAATTAAGTCCATTTT 1040  
 DB 961 GTCTGTCTAAACGAAATGGGTGGGAAAAAGTATGGTCCAAATTAAGTCCATTTT 1020  
 QY 1041 GCCATTAATGGCAATCTTGCCTTTGTTATTTTGGTGCAGTGTCTTCTGCTTAATCAT 1100  
 DB 1021 GCCATTAATGGCAATCTTGCCTTTGTTATTTTGGTGCAGTGTCTTCTGCTTAATCAT 1080  
 QY 1101 TTGCTTTTGGCATCTGTGTTTATTTTACTTTGACACACATGCAAGTTTACATCTGCTT 1160  
 DB 1081 TTGCTTTTGGCATCTGTGTTTATTTTACTTTGACACACATGCAAGTTTACATCTGCTT 1140  
 QY 1161 AACTACTCTTCCGAGGTAAATCCAAATTTATTTTGAACATCCGCTTAAGGGGCCCATCT 1220  
 DB 1141 AACTACTCTTCCGAGGTAAATCCAAATTTATTTTGAACATCCGCTTAAGGGGCCCATCT 1200  
 QY 1221 CTCTCACTCTTCTTCTAGTCAATATTCAGCAAAATTTATTTTGAAGCCCTTACTGAGG 1280  
 DB 1201 CTCTCACTCTTCTTCTAGTCAATATTCAGCAAAATTTATTTTGAAGCCCTTACTGAGG 1260  
 QY 1281 CAAATCATTTGTACTGGATTAATGAAGAAAAATAGATAATTCCTTATTCAGTAAATGCTTA 1340  
 DB 1261 CAAATCATTTGTACTGGATTAATGAAGAAAAATAGATAATTCCTTATTCAGTAAATGCTTA 1320  
 QY 1341 CTGAGCACAATCTAGTAATCATTTACAGTAATGGCCCATGTTTGTGTTGAGGTGTGTA 1400  
 DB 1321 CTGAGCACAATCTAGTAATCATTTACAGTAATGGCCCATGTTTGTGTTGAGGTGTGTA 1380  
 QY 1401 TTGATAACATATTTTACACCATTCGATCAATGATTAATTAAGACAAATATATACATCA 1460  
 DB 1381 TTGATAACATATTTTACACCATTCGATCAATGATTAATTAAGACAAATATATACATCA 1440

QY 1461 AGATTAAGTAATTTGTTGTTATCTGCATTTTAAAGATCCAGATTTGATCAATTAAT 1520  
 DB 1441 AGATTAAGTAATTTGTTGTTATCTGCATTTTAAAGATCCAGATTTGATCAATTAAT 1500  
 QY 1521 TATTAATTAAGAAAAATGATTTAATCTGTAATTAATGATGTTTATTTGCACTGTA 1580  
 DB 1501 TATTAATTAAGAAAAATGATTTAATCTGTAATTAATGATGTTTATTTGCACTGTA 1560  
 QY 1581 ATATCTAGATTTTATTAATTTGTTTAACTGCTCCACCAACATGCTAGATTAATA 1640  
 DB 1561 ATATCTAGATTTTATTAATTTGTTTAACTGCTCCACCAACATGCTAGATTAATA 1620  
 QY 1641 CCCCCAAATTAAGTATTTAACTTTGATAGTATTAAGAGACTGGGTGCTAATTAATAG 1700  
 DB 1621 CCCCCAAATTAAGTATTTAACTTTGATAGTATTAAGAGACTGGGTGCTAATTAATAG 1680  
 QY 1701 ATTAATTTGAGGCAAGAGAGCTGTTATCTTAAGTATTAATTTGATGTTCTGTAATGA 1760  
 DB 1681 ATTAATTTGAGGCAAGAGAGCTGTTATCTTAAGTATTAATTTGATGTTCTGTAATGA 1740  
 QY 1761 GAAATGTTGACCAATTAATTAATTTTATAGTATTAATTAATTAATTTTATAGGAGCATG 1820  
 DB 1741 GAAATGTTGACCAATTAATTAATTTTATAGTATTAATTAATTAATTTTATAGGAGCATG 1800  
 QY 1821 TTCTGTATTAAGCAATTAATTAATTTTATAGTATCA 1857  
 DB 1801 TTCTGTATTAAGCAATTAATTAATTTTATAGTATCA 1837

RESULT 4  
 AAH14124  
 ID AAH14124 standard; cDNA; 1836 BP.

AC AAH14124;  
 DT 26-JUN-2001 (first entry)  
 XX Human cDNA sequence SEQ ID NO:11319.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 OS Homo sapiens.

PN EP1074617-A2.  
 XX 07-FEB-2001.

PD 28-JUL-2000; 2000EP-00116126.  
 PF 29-AUG-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INSTR.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 DR WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.

PS Claim 8; SEQ ID NO 11319; 2537bp + Sequence Listing; English.

CC The present invention describes primer sets for synthesizing 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602

CC nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of the  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AHH3162 to AHH1628 and  
CC AHH1363 to AHH1874 represent human cDNA sequences; AHH92446 to AHH95893  
CC represent human amino acid sequences; and AHH13629 to AHH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention.

XX Sequence 1836 BP; 562 A; 330 C; 386 G; 558 T; 0 U; 0 Other;

Query Match 79.2%; Score 1483; DB 4; Length 1836;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1833; Conservative 0; Mismatches 1; Indels 3; Gaps 2;

```

QY 22 GAGGCTGCGCCGCTGAGACCGCGAGCGTTGCTGAGCCCGTTAGTGGCCCGCGCAGA 81
Db 1 GAGGCTGCGCCGCTGAGACCGCGAGCGTTGCTGAGCCCGTTAGTGGCCCGCGCAGA 60
QY 82 CACGCGCGCGCATGTCCTGCTACCTGCGTCCCGCCCAACAGCTCTGTTGCTGAGAAC 141
Db 61 CACGCGCGCGCATGTCCTGCTACCTGCGTCCCGCCCAACAGCTCTGTTGCTGAGAAC 120
QY 142 GTGGCCGACGACACACGAGTGTGAAGACTTGGCGGCTGAATTTGCTGTTATGCTCTATA 201
Db 121 GTGGCCGACGACACACGAGTGTGAAGACTTGGCGGCTGAATTTGCTGTTATGCTCTATA 180
QY 202 GTTATGTTGTTATGTTCCCTGATTTTCTACACTGCGCGCTGCAAGAGATTGCTTATGTT 261
Db 181 GTTATGTTGTTATGTTCCCTGATTTTCTACACTGCGCGCTGCAAGAGATTGCTTATGTT 240
QY 262 CAATTGAGATGTTCTGATGCTGAAGACGCTTACATTAATTGGAAGAAAGTGAT 321
Db 241 CAATTGAGATGTTCTGATGCTGAAGACGCTTACATTAATTGGAAGAAAGTGAT 300
QY 322 TGTGGACGCGAGATTGAATATACGTTTGCCAGGGGGATCGAAGACACCAATAGATG 381
Db 301 TGTGGACGCGAGATTGAATATACGTTTGCCAGGGGGATCGAAGACACCAATAGATG 360
QY 382 AAAGCGAAGAGAGAGAGATGTGACAGTCTTCAAGCTATGATGATTAGACAGATAC 441
Db 361 AAAGCGAAGAGAGAGAGATGTGACAGTCTTCAAGCTATGATGATTAGACAGATAC 420
QY 442 AGACCTTGAAGAGCGAAGTTATGAAGAGAGATCAAGAAGTCGGTCTTTGATTAC 501
Db 421 AGACCTTGAAGAGCGAAGTTATGAAGAGAGATCAAGAAGTCGGTCTTTGATTAC 480
QY 502 AACCTATGAAGATGCTATATGCTTGAAGACAGTACACGACTGGAAGACACAGCGGTAG- 560
Db 481 AACCTATGAAGATGCTATATGCTTGAAGACAGTACACGACTGGAAGACACAGCGGTAGC 540
QY 561 AGAAGCCATTCCGACAAATGATAGACCAAACTGACAGTGAATATCCAGTACAGTCTGCT 620
Db 541 AGAAGCCATTCCGACAAATGATAGACCAAACTGACAGTGAATATCCAGTACAGTCTGCT 600
QY 621 TACTACACTTCAAGAAAGATCTGAAGCGGAAAAAGAACCAAGAAAGAGGGCAGTTCAGAGG 680
Db 601 TACTACACTTCAAGAAAGATCTGAAGCGGAAAAAGAACCAAGAAAGAGGGCAGTTCAGAGG 660
QY 681 ACCAAAAGGGTGGTGAAGGTGCTGACATGATATCTGATAGATATTTTGACTCTGTT 740
Db 661 ACCAAAAGGGTGGTGAAGGTGCTGACATGATATCTGATAGATATTTTGACTCTGTT 720
```

```

QY 741 CTGAAGATTAAGAAAGATGTTATCGAAAACCTACATGGAATTAATTGAAGTCCTTCAAGTT 800
Db 721 CTGAAGATTAAGAAAGATGTTATCGAAAACCTACATGGAATTAATTGAAGTCCTTCAAGTT 780
QY 801 TGAAGTAAAGCATTTTGAAGCAAAATTAAGAAATTCACCTTTGACTTGTGGAACCTAA 860
Db 781 TGAAGTAAAGCATTTTGAAGCAAAATTAAGAAATTCACCTTTGACTTGTGGAACCTAA 840
QY 861 TCCCTAAATTAAGATAGTTTATTTGATTCAGGGGTAACAGGTCCTAATTAATTAATTG 920
Db 841 TCCCTAAATTAAGATAGTTTATTTGATTCAGGGGTAACAGGTCCTAATTAATTAATTG 900
QY 921 GAAACTAGATAGTGTGATATCAAGGAAGACAGCCATAGTCTCTTACAGTGGCTCTGTG 980
Db 901 GAAACTAGATAGTGTGATATCAAGGAAGACAGCCATAGTCTCTTACAGTGGCTCTGTG 960
QY 981 GTCTGTCTCAAACTGAATTTGGGTGGGAAAAGATAGTGTCCATATTAAGTTCCATTTT 1040
Db 961 GTCTGTCTCAAACTGAATTTGGGTGGGAAAAGATAGTGTCCATATTAAGTTCCATTTT 1020
QY 1041 GCCATTAATTTGGCAATCTTGCTTGTATTTTGTGGCGAGGTGTTCTGCTTAATCAT 1100
Db 1021 GCCATTAATTTGGCAATCTTGCTTGTATTTTGTGGCGAGGTGTTCTGCTTAATCAT 1080
QY 1101 TTGCTTTGTTGGCATCTGTGTTTATTTACTTGTACACACATGACATGATCATCTGTCT 1160
Db 1081 TTGCTTTGTTGGCATCTGTG--TATTTACTTGTACACACATGACATGATCATCTGTCT 1138
QY 1161 AACTACTCTTCCGAGGTAATTCOAATTAATTTGACATCCAGCTAAGAGGCGCCATCT 1220
Db 1139 AACTACTCTTCCGAGGTAATTCOAATTAATTTGACATCCAGCTAAGAGGCGCCATCT 1198
QY 1221 CTCTGACCTCTTCCATGACATTAATTTGCAAAATTAATTTAGAGCCCTTACGTGAGG 1280
Db 1199 CTCTGACCTCTTCCATGACATTAATTTGCAAAATTAATTTAGAGCCCTTACGTGAGG 1258
QY 1281 CAATTCATTTGACTGATTAATTTGAAGAAATTAATTTCCCTTAATTCAGTAAATGTCTA 1340
Db 1259 CAATTCATTTGACTGATTAATTTGAAGAAATTAATTTCCCTTAATTCAGTAAATGTCTA 1318
QY 1341 CTGAGCAATCTAGTGAATCATTAACGATAGGCTCATTTGTTTGTAGAGTGTGTTA 1400
Db 1319 CTGAGCAATCTAGTGAATCATTAACGATAGGCTCATTTGTTTGTAGAGTGTGTTA 1378
QY 1401 TTCAATCAATTAATTTTACACCATTCGATCAATGATTAATTAAGAACACAAATACATCA 1460
Db 1379 TTCAATCAATTAATTTTACACCATTCGATCAATGATTAATTAAGAACACAAATACATCA 1438
QY 1461 AGGATTAAGTAATTTGTGTTATCTGCATTTAAAGATATCCAGTATTTGATCACATTA 1520
Db 1439 AGGATTAAGTAATTTGTGTTATCTGCATTTAAAGATATCCAGTATTTGATCACATTA 1498
QY 1521 TATAAATTAATTAAGAAATTAATTTATCTGTAATTAACCTGTTTATTTGGAGTGTGCTGA 1580
Db 1499 TATAAATTAATTAAGAAATTAATTTATCTGTAATTAACCTGTTTATTTGGAGTGTGCTGA 1558
QY 1581 ATATCTAGATTAATTAATTAATTTTACTCTGCTCAACCAACACATGCTAGATATATA 1640
Db 1559 ATATCTAGATTAATTAATTAATTTTACTCTGCTCAACCAACACATGCTAGATATATA 1618
QY 1641 CCCCCAAATTAATTAATTTTACTCTGATTAAGTATTAAGAGACGTGGTGTCTAATATTA 1700
Db 1619 CCCCCAAATTAATTAATTTTACTCTGATTAAGTATTAAGAGACGTGGTGTCTAATATTA 1678
QY 1701 ATTAATTTGAGGACACAGAGAGCTGTATCTCTAATCTAATTTAGATGTTCTGTAATTA 1760
Db 1679 ATTAATTTGAGGACACAGAGAGCTGTATCTCTAATCTAATTTAGATGTTCTGTAATTA 1738
QY 1761 GAAAATGTTCAACCAATTAATCTTTTATGATTAATTAATTAATTAATTAATTAATTA 1820
Db 1739 GAAAATGTTCAACCAATTAATCTTTTATGATTAATTAATTAATTAATTAATTAATTA 1798
QY 1821 TTCTGTATTAAGCAATTAATTAATTTATTAATTAATTAATTAATTAATTAATTA 1857
```

Db	1799	TTCTGTGATAGCGAATAAATACTTTATAGATCA	1835
RESULT 5			
ID	AAZ51258		
AC	AAZ51258 standard; cDNA; 1326 BP.		
XX			
XX	AAZ51258;		
DT	06-JUN-2000 (first entry)		
DE			
XX			
XX	Human RNA-associated protein-9 (RNAAP-9) encoding cDNA.		
KW	RNA-associated protein; RNAAP; human; clone 934406; cytosolic; immunosuppressive; antiinflammatory; keratolytic; neuroprotective; antiarteriosclerotic; hepatocarcinoma; antipsoriatic; vinnicide; anti-HIV; antiatheric; antithrombotic; antiarthritic; ophthalmological; autoimmune; antimicrobial; cell proliferative disorder; inflammation; cirrhosis; actinic keratosis; burns; arteriosclerosis; atherosclerosis; hepatitis; myelofibrosis; primary thrombocytopenia; psoriasis; cancer; mixed connective tissue disease; MCTD; HIV; uveitis; Crohn's disease; allergy; rheumatoid arthritis; parasitic infection; ss.		
OS	Homo sapiens.		
XX			
XX	Key	Location/Qualifiers	
XX	CDS	222..773	
FT		/*tag= a	
FT		/product= "Human RNA-associated protein-9"	
FT		/note= "Derived from CERVO101 library"	
XX			
XX	MO20001171-A2.		
XX	02-MAR-2000.		
PD			
PF	20-AUG-1999; 99WO-US019361.		
XX			
XX	21-AUG-1998; 98US-0097550P.		
ER	12-JAN-1999; 99US-0115639P.		
PA	(INCY-) INCYTE PHARM INC.		
XX			
PI	Hillman JL, Yue H, Tang YT, Corley NC, Guegler KJ, Gorgone GA, Patterson C, Baughn MS, Lal P, Bandman O, Reddy R, Azimzai Y, Shih LL, Yang J, Lu DAM;		
XX			
DR	WPI; 2000-237651/20.		
XX	P-PSDB; AAY70228.		
PT	Human RNA-associated proteins useful in diagnosing, treating and preventing cell proliferative, autoimmune, inflammatory and infectious disorders.		
PS			
XX	Claim 9; Page 112; 123pp; English.		
XX			
CC	The present sequence is the cDNA encoding human RNA-associated protein-9 (RNAAP-9), identified in Incyte clone 934406, derived from CERVO101 library. It is expressed in reproductive, nervous, cardiovascular and haematopoietic/immune tissues. It has cytosolic, immunosuppressive, antiinflammatory, antiarteriosclerotic, hepatocarcinoma, keratolytic, neuroprotective, antipsoriatic, anti-HIV, antiatheric, antithrombotic, vinnicide, antiarthritic, ophthalmological and antimicrobial activity. RNAAP antibodies are useful for diagnosis of diseases associated with altered expression or activity of RNAAP. It is used to treat cell proliferative, autoimmune, inflammatory and infectious disorders, like actinic keratosis, burns, arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, myelofibrosis, mixed connective tissue disease (MCTD), psoriasis, primary thrombocytopenia and cancer, HIV, allergies, rheumatoid arthritis, uveitis, Crohn's disease, and bacterial, viral and parasitic infections		
XX			
XX	Sequence 1326 BP; 370 A; 261 C; 337 G; 357 T; 0 U; 1 Other;		

Query Match	50.2%	Score 940:	DB 3;	Length 1326;	
Best Local Similarity	99.8%;	Pred. No. 0;			
Matches 1180;	Conservative	0;	Mismatches	0;	Indels 2;
					Gaps 2;
QY	1	GGCCCTCTAGCTGTGTGTGTCTGAGGCTTCGCGCGCCCTGAGCGCGGACGGTTTGGCTGACCC	60		
Db	129	GCCCTCTAGCTGTGTGTGTCTGAGGCTTCGCGCGCCCTGAGCGCGGACGGTTTGGCTGAGCC	188		
QY	61	CGTTAGTGGCCCGGCGCGAGACAGCGCGCGCCATGTCCCGCTACCTGCTCCGCCAAC	120		
Db	189	CGTTAGTGGCCCGGCGCGAGAGAGCGCGCGCCATGTCCCGCTACCTGCTCCGCCAAC	248		
QY	121	ACGCTCTCGTTCGTCAGGAACGTGGCCGACGACACACAGTCTGAAAGATTTTGGGGGTGA	180		
Db	249	ACGCTCTCGTTCGTCAGGAACGTGGCCGACGACACACAGTCTGAAAGATTTTGGGGGTGA	308		
QY	181	TTTGGTCGTTATNGCTCCATAGTTGATGTGATGTTCCACTTGAATTTCTACACTGCGCGT	240		
Db	309	TTTGGTCGTTATNGCTCCATAGTTGATGTGATGTTCCACTTGAATTTCTACACTGCGCGT	368		
QY	241	CCAAAGAGATTTCCTTATGTTCAATTTGAGGATGTTGCTGATGCTGAAGACGCTTTTACAT	300		
Db	369	CCAAAGAGATTTCCTTATGTTCAATTTGAGGATGTTGCTGATGCTGAAGACGCTTTTACAT	428		
QY	301	AATTTGACAGAAAGTGGATTTGTGACCGGACAGTTGAAATACACTTTTCCGACGGGAT	360		
Db	429	AATTTGACAGAAAGTGGATTTGTGACCGGACAGTTGAAATACACTTTTCCGACGGGAT	488		
QY	361	CGAAAGACACCAAAATCAGATGAAAGCCAGAAAGGAGAGATGTGTACAGTTCTTCACGC	420		
Db	489	CGAAAGACACCAAAATCAGATGAAAGCCAGAAAGGAGAGATGTGTACAGTTCTTCACGC	548		
QY	421	TATGATGATTATGACAGATACAGACGTTCTTGAAAGCCGAGTTATGAAAGAGAGATCA	480		
Db	549	TATGATGATTATGACAGATACAGACGTTCTTGAAAGCCGAGTTATGAAAGAGAGATCA	608		
QY	481	AGAAAGTCGGTCTTTTGGATTACACTATAGAAAGTCGTATCTCATAGAAACAGTAGACCG	540		
Db	609	AGAAAGTCGGTCTTTTGGATTACACTATAGAAAGTCGTATCTCATAGAAACAGTAGACCG	668		
QY	541	ACTGGAAGACACGCGCGCTAG-AGAAAGCCATTCCGACATGATAGACCCAACTCGACGCTGG	599		
Db	669	ACTGGAAGACACGCGCGCTAGACGAAAGCCATTCCGACATGATAGACCCAACTCGACGCTGG	728		
QY	600	AATACCCAGTACAGTTCTGCTTACTACACTTCAGAAAGATCTGAAACCGGAAAAAGAAC	659		
Db	729	AATACCCAGTACAGTTCTGCTTACTACACTTCAGAAAGATCTGAAACCGGAAAAAGAAC	788		
QY	660	CAAAAGAGGGCAGTTCAAGCGACCAAAAGGGGTGGAAAGGTGCTGCAGTAGTAATCTG	719		
Db	789	CAAAAGAGGGCAGTTCAAGCGACCAAAAGGGGTGGAAAGGTGCTGCAGTAGTAATCTG	848		
QY	720	TACGAAATATTTGACTCTGCTCTGAAAAAGATAAAGAGATGTTATCGAAAACTACATGGAA	779		
Db	849	TACGAAATATTTGACTCTGCTCTGAAAAAGATAAAGAGATGTTATCGAAAACTACATGGAA	908		
QY	780	TAATTTGAAGTCCCTTCAAAGTTGAAAGTAGCATTTTATGACAAATATAAAGAAATTTCAA	839		
Db	909	TAATTTGAAGTCCCTTCAAAGTTGAAAGTAGCATTTTATGACAAATATAAAGAAATTTCAA	968		
QY	840	CTTTGTACTTGGGAAACTAATCCCTAAATATGAAATGGTTATATGATTCACAGGGTAA	899		
Db	969	CTTTGTACTTGGGAAACTAATCCCTAAATATGAAATGGTTATATGATTCACAGGGTAA	1028		
QY	900	CAGGTCCATATAAATATATGGAACCTAGATGTCTGAATATCAAGAGAGACAGCCATAG	959		
Db	1029	CAGGTCCATATAAATATATGGAACCTAGATGTCTGAATATCAAGAGAGACAGCCATAG	1088		
QY	960	TCTCTTACAGTGCCTCTGTTGCTGTGTCTCAAACTGAAATTTGGGTGGAAAAAGTATGTC	1019		
Db	1089	TCTCTTACAGTGCCTCTGTTGCTGTGTCTCAAACTGAAATTTGGGTGGAAAAAGTATGTC	1148		



QY 1020 CAATATAAGTTCATTTTGGCATTAAT-GGCAATCTGCGCTTGTATTTTGGTG 1078  
DB 1149 CAATATAAGTTCATTTTGGCATTAATGGGCAAACTTGCCTTGTATTTTGGTG 1208  
QY 1079 CCAGTGTTCCTGCTTAATTCATTTGCTTGTGTCATCTGCTTGTATTTTACTTATACACC 1138  
DB 1209 CCAGTGTTCCTGCTTAATTCATTTGCTTGTGTCATCTGCTTGTATTTTACTTATACACC 1268  
QY 1139 ACATGCACTTACATCTGCTTAACTACTCCTTCCAGGTAA 1180  
DB 1269 ACATGCACTTACATCTGCTTAACTACTCCTTCCAGGTAA 1310

RESULT 6  
ADA52465  
ID ADA52465 standard, cDNA, 2473 BP.  
AC ADA52465;  
AT 20-NOV-2003 (first entry)  
DT 20-NOV-2003 (first entry)  
DE Human coding sequence, SEQ ID 33.  
XX  
XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;  
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;  
KW inflammatory disease; osteoporosis; neurological disease; gene; ss.  
XX Homo sapiens.  
OS  
XX  
XX EP1293569-A2.  
XX  
XX PD 19-MAR-2003.  
XX  
XX PF 21-MAR-2002; 2002EP-00006586.  
XX  
XX PR 14-SEP-2001; 2001JP-00328381.  
XX PR 24-JAN-2002; 2002US-0350435P.  
XX  
XX PA (HELI-) HELIX RES INST.  
XX (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
XX PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,  
XX PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
XX PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;  
XX  
XX DR MPI: 2003-395539/38.  
XX DR P-PSDB; ADA54104.  
XX  
XX PS Claim 1; SEQ ID NO 33; 2055bp; English.  
XX  
XX CC The present invention relates to novel human secretory or membrane  
XX CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
XX CC ADA54071). The coding sequences are useful in the gene therapy of  
XX CC diseases caused by abnormalities of the proteins, e.g. cancer,  
XX CC inflammatory diseases, osteoporosis or neurological disease.  
XX  
XX SQ Sequence 2473 BP; 791 A; 395 C; 459 G; 828 T; 0 U; 0 Other;

Query Match 42.0%; Score 786; DB 7; Length 2473;  
Best Local Similarity 99.9%; Pred. No. 2,1e-300;  
Matches 836; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 582 AGACCAACTGCACTGGAATACCCAGTACAGTTCCTTACTTACACTTCAAGAAAGATC 641  
DB 1388 AGACCAACTGCACTGGAATACCCAGTACAGTTCCTTACTTACACTTCAAGAAAGATC 1447  
QY 642 TGAAGCGGAAAAAGAACCAAGAAAGGAGGAGTTCAGAGCACAAGAGGGTGGTGAAGGT 701  
DB 1448 TGAAGCGGAAAAAGAACCAAGAAAGGAGGAGTTCAGAGCACAAGAGGGTGGTGAAGGT 1507

QY 702 GCTGAGATATGAAATACCTAGCAATATTTTGAAGCTGCTGTAAGAAAGATGTT 761  
DB 1508 GCTGAGATATGAAATACCTAGCAATATTTTGAAGCTGCTGTAAGAAAGATGTT 1567  
QY 762 ATCGAAACTACATGAAATTAATGGAAGTCCCTCAAGTTTGAAGTAAGCAATTTAGAGC 821  
DB 1568 ATCGAAACTACATGAAATTAATGGAAGTCCCTCAAGTTTGAAGTAAGCAATTTAGAGC 1627  
QY 822 AAATTAAGAAATTCAACTTTGTAAGTTGGAAGCTATCCCTTAATATGATATGTTT 881  
DB 1628 AAATTAAGAAATTCAACTTTGTAAGTTGGAAGCTATCCCTTAATATGATATGTTT 1687  
QY 882 ATATTGATTCATGGGTAAAGGTCATATTAATTAATTAATTAATTAATTAATTAATTAAT 941  
DB 1688 ATATTGATTCATGGGTAAAGGTCATATTAATTAATTAATTAATTAATTAATTAATTAAT 1747  
QY 942 CAAGGAAGACAGCCATAGTCTCTTACAGTGCCTCTGTTGCTGCTCAAACTGAATTGG 1001  
DB 1748 CAAGGAAGACAGCCATAGTCTCTTACAGTGCCTCTGTTGCTGCTCAAACTGAATTGG 1807  
QY 1002 GTGGAAAAAGTATGCTCCATATTAAGTTTCATTTTGCATATTTGGCAAAATCTTGC 1061  
DB 1808 GTGGAAAAAGTATGCTCCATATTAAGTTTCATTTTGCATATTTGGCAAAATCTTGC 1867  
QY 1062 CTTTGTATTTTGGTGGCAGTGTTCCTGCTTAAATCAATTTGCTTGGTGGCATCTGTGT 1121  
DB 1868 CTTTGTATTTTGGTGGCAGTGTTCCTGCTTAAATCAATTTGCTTGGTGGCATCTGTGT 1927  
QY 1122 TTATTTACTTGTACACACATGACAGTTTACATCTGCTTAACTACTCTCTCCAGGTAA 1181  
DB 1928 TTATTTACTTGTACACACATGACAGTTTACATCTGCTTAACTACTCTCTCCAGGTAA 1987  
QY 1182 TTCCAAATTAATTTTACATCCAGCTAAGAGGCGCATCTCTCTACCCCTTCCCTAGTC 1241  
DB 1988 TTCCAAATTAATTTTACATCCAGCTAAGAGGCGCATCTCTCTACCCCTTCCCTAGTC 2047  
QY 1242 AGTATATTCAGCAATATTTATTTAGCCCTTACCTGAGGCAATCATTTGTAATAT 1301  
DB 2048 AGTATATTCAGCAATATTTATTTAGCCCTTACCTGAGGCAATCATTTGTAATAT 2107  
QY 1302 TGAGAAAAATAGATATATTCCTTATTCAGTAAATATGCTACAGCAACATCTAGGAATC 1361  
DB 2108 TGAGAAAAATAGATATATTCCTTATTCAGTAAATATGCTACAGCAACATCTAGGAATC 2167  
QY 1362 ATTACAGTATGSCCTCATTTGTTTGTGAGGTGTGTTATTCATACAAATATTTTAC 1418  
DB 2168 ATTACAGTATGSCCTCATTTGTTTGTGAGGTGTGTTATTCATACAAATATTTTAC 2224

RESULT 7  
ABV25632  
ID ABV25632 standard, cDNA, 1382 BP.  
AC ABV25632;  
AT 16-SEP-2002 (first entry)  
DT 16-SEP-2002 (first entry)  
DE Human prostate expression marker CDNA 25623.  
XX  
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
XX KW pharmacogenomic marker; gene; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200160860-A2.  
XX  
XX PD 23-AUG-2001.  
XX  
XX PF 20-FEB-2001; 2001WO-US005171.  
XX  
XX PR 17-FEB-2000; 2000US-0183319P.  
XX PR 16-MAR-2000; 2000US-0189862P.

PR	25-MAY-2000; 2000US-0207454P.
PR	09-JUN-2000; 2000US-0211314P.
PR	18-JUN-2000; 2000US-0219007P.
PR	13-DEC-2000; 2000US-0255281P.
XX	
XX	
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX	
PI	Schlegel R, Endege WO, Monahan JE;
XX	
DR	WPI; 2001-662795/76.
XX	
PT	Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.
XX	
PS	Claim 1; Page 5128-5129; 11750pp; English.
CC	The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
CC	
CC	
CC	
CC	
CC	
SO	Sequence 1382 BP; 462 A; 282 C; 285 G; 340 T; 0 U; 13 Other;
Query Match	37.9%; Score 709; DB 5; Length 1382;
Best Local Similarity	100.0%; Pred. No. 5,5e-270;
Matches 709; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Dy	561 AGAAGCATTCGACATGATGATGACCAAACTGCAGTGGAAATACCAGTAGACTTGCT 620
Db	230 AGAAGCATTCGACATGATGATGACCAAACTGCAGTGGAAATACCAGTAGACTTGCT 289
Dy	621 TACATCAGCTTCAAGAAAGATCTGAAAAGCGGAAAAAGAACCAAAGAAGGCAGTTCAACG 680
Db	290 TACATCAGCTTCAAGAAAGATCTGAAAAGCGGAAAAAGAACCAAAGAAGGCAGTTCAACG 349
Dy	681 ACCAAAGGGGTGGTGAAGGTGCTGCAATGAAATACTGTACGAAATATTGGAGCTCGGT 740
Db	350 ACCAAAGGGGTGGTGAAGGTGCTGCAATGAAATACTGTACGAAATATTGGAGCTCGGT 409
Dy	741 CTGAAAAGATTAAGAAATGTTATCGAAAACCTACATGAAATATTGAATGCCCTCAAGTT 800
Db	410 CTGAAAAGATTAAGAAATGTTATCGAAAACCTACATGAAATATTGAATGCCCTCAAGTT 469
Dy	801 TGAAAGTAGAATTTTAGACAATAATTAAGAAATTCACCTTTGTACTTGTGAAACTCA 860
Db	470 TGAAAGTAGAATTTTAGACAATAATTAAGAAATTCACCTTTGTACTTGTGAAACTCA 529
Dy	861 TCCCTAAATATGAAATAGATTATATTTGATCATGGGTACAGGTCACATAAATTAATT 920
Db	530 TCCCTAAATATGAAATAGATTATATTTGATCATGGGTACAGGTCACATAAATTAATT 589
Dy	921 GAAACATGAGATGTCGAAATATCAAGAAAGACAGCATAGTCTTACAGTCCCTCTGTG 980
Db	590 GAAACATGAGATGTCGAAATATCAAGAAAGACAGCATAGTCTTACAGTCCCTCTGTG 649
Dy	981 GTCTGTCTCAAACTGAATTGGGTGGGAAAAGGTATGTCAAATATTAAGTTCCAATTTT 1040
Db	650 GTCTGTCTCAAACTGAATTGGGTGGGAAAAGGTATGTCAAATATTAAGTTCCAATTTT 709
Dy	1041 GCCATTAATGGAACAATCTGCCTTTGTTATTTTGGGACAGTGTTTTCSCTTAATCAT 1100
Db	710 GCCATTAATGGAACAATCTGCCTTTGTTATTTTGGGACAGTGTTTTCTCTTAATCAT 769
Dy	1101 TTGCTTTGTTGGACCTGTGTTTATTTTATCTTGTAACCAACATGCAAGTTACATGTCTTT 1160

Db	770	TTGCTTTGGTGGCACTGTGTATTATTAATTGTAACATCTGTCTT	829
Qy	1161	AACTACTCCTTCCAGGTAAATTCCAATATATTGACATCCAGTAAAGAGGCCATCT	1222
Db	830	AACTACTCCTTCCAGGTAAATTCCAATATATTGATCATCCAGTAAAGAGGCCATCT	889
Qy	1221	CTTCCAGCTCTTTCCTTGTAGTATATTCGACAAATATTTATGAGCC	1269
Db	890	CTTCCAGCTCTTTCCTTGTAGTATATTCGACAAATATTTATGAGCC	938
RESULT 8			
ID	AAH14349	standard; cDNA, 2247 BP.	
XX	AAH14349;		
AC	26-JUN-2001	(first entry)	
DT			
XX			
DE	Human cDNA sequence SEQ ID NO:11739.		
XX			
XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.		
XX	Os		
XX	Homo sapiens.		
XX	EP1074617-A2.		
PD	07-FEB-2001.		
XX			
PF	28-JUL-2000; 2000EP-00116126.		
XX			
PR	29-JUL-1999; 99JP-00248036.		
PR	27-AUG-1999; 99JP-00300253.		
PR	11-JAN-2000; 2000JP-00118776.		
PR	02-MAY-2000; 2000JP-00183767.		
PR	09-JUN-2000; 2000JP-00241899.		
XX			
PA	(HELI-) HELIX RES INST.		
XX			
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
XX			
XX	WPI; 2001-318749/34.		
PT			
PT	Primer sets for synthesizing polynucleotides, particularly the 5602 full-		
PT	length cDNAs defined in the specification, and for the detection and/or		
PT	diagnosis of the abnormality of the proteins encoded by the full-length		
PT	cDNAs.		
XX			
XX	Claim 8; SEQ ID NO 11739; 2537bp + Sequence Listing; English.		
PS			
XX			
XX	The present invention describes primer sets for synthesizing 5602 full-		
CC	length cDNAs defined in the specification. Where a primer set comprises:		
CC	(a) an oligo-dT primer and an oligonucleotide complementary to the		
CC	complementary strand of a polynucleotide which comprises one of the 5602		
CC	nucleotide sequences defined in the specification, where the		
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination		
CC	of an oligonucleotide comprising a sequence complementary to the		
CC	complementary strand of a polynucleotide which comprises a 5'-end		
CC	sequence and an oligonucleotide comprising a sequence complementary to a		
CC	polynucleotide which comprises a 3'-end sequence, where the		
CC	oligonucleotide comprises at least 15 nucleotides and the combination of		
CC	the 5'-end sequence/3'-end sequence is selected from those defined in the		
CC	specification. The primer sets can be used in antisense therapy and in		
CC	gene therapy. The primers are useful for synthesizing polynucleotides,		
CC	particularly full-length cDNAs. The primers are also useful for the		
CC	detection and/or diagnosis of the abnormality of the proteins encoded by		
CC	the full-length cDNAs. The primers allow obtaining of the full-length		
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and		
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893		
CC	represent human amino acid sequences; and AAH13629 to AAH13632 represent		
CC	oligonucleotides, all of which are used in the exemplification of the		

CC present invention

XX Sequence 2247 BP, 669 A, 348 C, 389 G, 841 T, 0 U, 0 Other;

SQ Query Match 36.0%; Score 674; DB 4; Length 2247;

Best Local Similarity 99.9%; Pred. No. 3,4e-256;

Matches 724; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1133 TACACCAATGAGCTTACATCTGCTTACACCTCCAGGTAATTCATTTA 1192

Db 1 TACACCAATGAGCTTACATCTGCTTACACCTCCAGGTAATTCATTTA 60

QY 1193 TTGACATCCAGCTAGAGGGCCCATCTCTCACTTCCAGTCAATATTAG 1252

Db 61 TTGACATCCAGCTAGAGGGCCCATCTCTCACTTCCAGTCAATATTAG 120

QY 1253 CAATATTTATTTAGCCCTTACTGCTGGGCAATTCATTTAGTAATGAAATA 1312

Db 121 CAATATTTATTTAGCCCTTACTGCTGGGCAATTCATTTAGTAATGAAATA 180

QY 1313 GATATTTCCCTATTCAGTAAATGCTCTAGCAATCTAGTAATTCAGTATG 1372

Db 181 GATATTTCCCTATTCAGTAAATGCTCTAGCAATCTAGTAATTCAGTATG 240

QY 1373 GCTCATTTGTTTGTGAGGTGTGTATTCATTAATTTTACACCATTCGATGA 1432

Db 241 GCTCATTTGTTTGTGAGGTGTGTATTCATTAATTTTACACCATTCGATGA 300

QY 1433 TGTATTTATAGAACCAATATTCAGTAAAGTAATTTGTGTTATTCGCAATT 1492

Db 301 TGTATTTATAGAACCAATATTCAGTAAAGTAATTTGTGTTATTCGCAATT 360

QY 1493 AAAAGTATCCAGTATTTAGTATCACTATTTAATTAAGTAAATTTAATCTGTA 1552

Db 361 AAAAGTATCCAGTATTTAGTATCACTATTTAATTAAGTAAATTTAATCTGTA 420

QY 1553 TAACTGTTTATTTGAGAGTGTATATATATATATATATATATATATATAT 1612

Db 421 TAACTGTTTATTTGAGAGTGTATATATATATATATATATATATATATAT 480

QY 1613 GCTTACCAACATCTGAGATATACCCCAAAATTAATTTAATTTGATAGG 1672

Db 481 GCTTACCAACATCTGAGATATACCCCAAAATTAATTTAATTTGATAGG 540

QY 1673 TATTAAGGAGCTGGGCTATTAATTTAGTATTTTGGAGGAGAGAGGCTTATCC 1732

Db 541 TATTAAGGAGCTGGGCTATTAATTTAGTATTTTGGAGGAGAGAGGCTTATCC 600

QY 1733 TAACTGATTTAGTATCTGCTATTTAGTATTTGAGGAGGAGGAGGAGGAG 1792

Db 601 TAACTGATTTAGTATCTGCTATTTAGTATTTGAGGAGGAGGAGGAGGAGG 660

QY 1793 TTTAATGATGACATTTTATAGGAGCATGTTGCTATAGGCAATTAATTTATAG 1852

Db 661 TTTAATGATGACATTTTATAGGAGCATGTTGCTATAGGCAATTAATTTATAG 720

QY 1853 TATCA 1857

Db 721 TATCA 725

RESULT 9

AAH06844 ID AAH06844 standard; cDNA; 702 BP.

XX AAH06844;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:3679.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

XX EPI074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Claim 1; SEQ ID NO 3679; 2537bp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-

XX length cDNAs defined in the specification. Where a primer set comprises:

XX (a) an oligo-dT primer and an oligonucleotide complementary to the

XX complementary strand of a polynucleotide which comprises one of the 5602

XX nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

XX polynucleotide which comprises a 3'-end sequence, where the

XX oligonucleotide comprises at least 15 nucleotides and the combination

XX of the 5'-end sequence/3'-end sequence is selected from those defined in the

XX specification. The primer sets can be used in antisense therapy and in

XX gene therapy. The primers are useful for synthesizing polynucleotides,

XX particularly full-length cDNAs. The primers are also useful for the

XX detection and/or diagnosis of the abnormality of the proteins encoded by

XX the full-length cDNAs. The primers allow obtaining of the full-length

XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

XX AAH13633 to AAH1742 represent human cDNA sequences; AAH92446 to AAH95893

XX represent human amino acid sequences; and AAH13629 to AAH13632 represent

XX oligonucleotides, all of which are used in the exemplification of the

XX present invention

XX Sequence 702 BP; 201 A; 159 C; 173 G; 166 T; 0 U; 3 Other;

XX Query Match 27.2%; Score 509; DB 4; Length 702;

XX Best Local Similarity 100.0%; Pred. No. 5.1e-191;

XX Matches 509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 GAGGCTGGCGCGCTGAGCGCGGAGCGGTTGCTGAGCCCGTTAGCGCGCGCGAGA 81

Db 1 GAGGCTGGCGCGCTGAGCGCGGAGCGGTTGCTGAGCCCGTTAGCGCGCGCGAGA 60

QY 82 CAGCGCGCGCGCATGCTCCGCTACCTGCGTCCCGCAACAGCGTCTGTGCTAGGAAC 141

Db 61 CAGCGCGCGCGCATGCTCCGCTACCTGCGTCCCGCAACAGCGTCTGTGCTAGGAAC 120

QY 142 GTGGCCGACGACACGAGTCTGAAGACTTGGCGGCGTGAATTTGCTGTTAGTCTTAA 201

Db 121 GTGGCCGACGACACGAGTCTGAAGACTTGGCGGCGTGAATTTGCTGTTAGTCTTAA 180

QY 202 GTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 261

Db 181 GTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240



```

XX 26-MAY-2000; 2000US-0207484P.
PR (CORI-) CORIXA CORP.
XX
XX Algate PA, Harlocker SL, Jones R;
XX WPI; 2002-122075/16.
XX
XX Composition for therapy and diagnosis of ovarian cancer comprising
XX polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
XX polypeptide, antibody specific to polypeptide or T cell expressing
XX polypeptide.
XX
XX Claim 1; SEQ ID NO 10911; 489pp; English.
XX
XX The present invention describes a composition (I) comprising: carriers
XX and immunostimulants; and a polypeptide (II) of a ovarian tumour
XX polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
XX from the 10912 nucleotide sequences as given in AB177023 to AB187934,
XX (III) encoding (II) having a sequence (S2), a T cell population of (II),
XX or antigen presenting cells that express (II). (I) has cytostatic
XX activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
XX detecting ovarian cancer in a patient's biological sample preferably
XX serum or ovarian tissue. The method comprises contacting a biological
XX sample from a patient with (IV), detecting the amount of polynucleotide
XX hybridising to (IV) and comparing the amount to a predetermined cutoff
XX value and thereby detecting ovarian cancer in the patient, where the
XX amount of polynucleotide hybridising to (IV) is detected preferably by
XX polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
XX useful for stimulating and/or expanding T cells specific for an ovarian
XX tumour protein comprising contacting T cells with (III) or (II). (III) is
XX useful in design and preparation of ribozyme molecules for inhibiting
XX expression of the tumour polypeptides and proteins in tumour cells; and
XX to isolate a full length gene from a suitable library e.g., a tumour cDNA
XX library using well known techniques
XX
XX Sequence 550 BP; 190 A; 89 C; 76 G; 194 T; 0 U; 1 Other;
XX
XX Query Match 24.6%; Score 460; DB 6; Length 550;
XX Best Local Similarity 99.8%; Pred. No. 1.1e-171;
XX Matches 510; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1347 ACAATCTGTAATGATTAAGTATGAGGCTCATGTTGGTTGGAGTGTATTCATA 1406
XX |||||||
XX 550 ACAATCTGTAATGATTAAGTATGAGGCTCATGTTGGTTGGAGTGTATTCATA 491
XX
XX 1407 ACAATATTTTACACCATTCGTATCAATGTAATTAAGAACAATATACGATCAAGATA 1466
XX |||||||
XX 490 ACAATATTTTACACCATTCGTATCAATGTAATTAAGAACAATATACGATCAAGATA 431
XX
XX 1467 AGTAATTTGGTGGTATTCGCCATTTAAAGATTCAGATTTTGCATTTATTA 1526
XX |||||||
XX 430 AGTAATTTGGTGGTATTCGCCATTTAAAGATTCAGATTTTGCATTTATTA 371
XX
XX 1527 TAATGAAAAAAGATTTAATCTGTAATAAAGCTGTTATGTCAGTGCCTAATATAC 1586
XX |||||||
XX 370 TAATGAAAAAAGATTTAATCTGTAATAAAGCTGTTATGTCAGTGCCTAATATAC 311
XX
XX 1587 TAGAGTTAATTAATTTGTTTACTGCTCCACCAACACATGCTAGATATACCCCA 1646
XX |||||||
XX 310 TAGAGTTAATTAATTTGTTTACTGCTCCACCAACACATGCTAGATATACCCCA 251
XX
XX 1647 AAATTAAGATTTAACTTTGATAGTATTAAGAGACTGGGTGCTATTAATTAAT 1706
XX |||||||
XX 250 AAATTAAGATTTAACTTTGATAGTATTAAGAGACTGGGTGCTATTAATTAAT 191
XX
XX 1707 TTGAGGCGAGAGAGAGGCTGATCCTAACTGATTTAGATGTTCTGTAATGAGAAAT 1766
XX |||||||
XX 190 TTGAGGCGAGAGAGAGGCTGATCCTAACTGATTTAGATGTTCTGTAATGAGAAAT 131
XX
XX 1767 GTTCCACCAATTTATACCTTTTATAGTATTAACATGTAATTTTATAGGGGACATGTTCTGT 1826
XX |||||||

```

```

DB 130 GTTCCACCAATTTATACCTTTTATAGTATTAACATGTAATTTTATAGGGGACATGTTCTGT 71
XX
XX 1827 GTATAGCGAATTAATAACTTTTATAGTATCA 1857
XX |||||||
XX 70 GTATAGCGAATTAATAACTTTTATAGTATCA 40
XX
XX RESULT 12
XX AAS79415
XX ID AAS79415 standard; cDNA; 716 BP.
XX AC AAS79415;
XX XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #15219.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Dymnac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX P-PSDB; ABG15228.
XX
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 1; SEQ ID NO 15219; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
XX coding sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 716 BP; 194 A; 163 C; 177 G; 182 T; 0 U; 0 Other;
XX
XX Query Match 23.2%; Score 435; DB 5; Length 716;
XX Best Local Similarity 100.0%; Pred. No. 8e-162;
XX Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



EN WO20016719-A1.  
XX  
XX 13-SEP-2001.  
XX  
XX 02-MAR-2001; 2001WO-JP001629.  
XX  
XX 07-MAR-2000; 2000JP-00159195.  
XX  
XX (CHIB-) CHIBA PREFECTURE.  
XX PA (HISM) HISAMITSU PHARM CO LTD.  
XX  
XX Nakagawara A;  
XX  
XX WPI; 2001-565584/63.  
XX  
XX Nucleic acids originating in gene expressed in human neuroblastoma,  
XX useful as probe or primer in diagnosing prognosis of human neuroblastoma,  
XX PT malignancy and susceptibility indicator or tumor marker for anti-cancer  
XX agents.  
XX  
XX Claim 1; Page 2100; 2979pp; Japanese.  
XX  
XX The invention relates to novel genes (AA193926-AA197963) expressed in  
XX human neuroblastoma. The nucleic acids are applicable as a probe or  
XX primer in diagnosing the prognosis of human neuroblastoma, malignancy and  
XX susceptibility indicators or tumour markers for anti-cancer agents. The  
XX gene information for diagnosing prognosis is related to factors similar  
XX to that for N-myc and Yrka genes  
XX  
SQ Sequence 737 BP; 213 A; 121 C; 128 G; 255 T; 0 U; 20 Other;  
  
Query Match 23.0%; Score 430; DB 4; Length 737;  
Best Local Similarity 99.8%; Pred. No. 7.5e-160;  
Matches 480; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1069 TATTTGGTGGCAGTGTTCGCTTAATCAATTTGCTTGGGATCGTGTATTTA 1128  
DB 48 TATTTGGTGGCAGTGTTCGCTTAATCAATTTGCTTGGGATCGTGTATTTA 107  
  
QY 1129 CTGTGACCCACATGCAATGCTTCACTACTCTCCAGGTAATTCAT 1188  
DB 108 CTGTGACCCACATGCAATGCTTCACTACTCTCCAGGTAATTCAT 167  
  
QY 1189 TATATTTGACATCCAGCTAAGAGGCCCATCTCTCACTCTTCTGATGATAT 1248  
DB 168 TATATTTGACATCCAGCTAAGAGGCCCATCTCTCACTCTTCTGATGATAT 227  
  
QY 1249 TCAGCAATATTTATTTAGAGCCCTTACTGTGGCAATGATGATGATGAGAA 1308  
DB 228 TCAGCAATATTTATTTAGAGCCCTTACTGTGGCAATGATGATGATGAGAA 287  
  
QY 1309 AATAGATATTTCCCTTATTCAGTAATGTCTTCTGAGCAATCTAGATCATTAAG 1368  
DB 288 AATAGATATTTCCCTTATTCAGTAATGTCTTCTGAGCAATCTAGATCATTAAG 347  
  
QY 1369 TATGGCCATGTTGTTGTTGAGGTGTATTCATTAACAATTTTACACCATTCGTA 1428  
DB 348 TATGGCCATGTTGTTGTTGAGGTGTATTCATTAACAATTTTACACCATTCGTA 407  
  
QY 1429 TCAATGTATATTAAGACACATATATAGATCAAGATAGTAATGTGTGTTATCTGCC 1488  
DB 408 TCAATGTATATTAAGACACATATATAGATCAAGATAGTAATGTGTGTTATCTGCC 467  
  
QY 1449 ATTTAAAGTATCCGATATTTGATCACTATATTAATAAATAAATAATGATTAATCT 1548  
DB 468 ATTTAAAGTATCCGATATTTGATCACTATATTAATAAATAAATAATGATTAATCT 527  
  
QY 1549 G 1549  
DB 528 G 528  
  
RESULT 15

ABV24425  
ID ABV24425 standard; cDNA; 2407 BP.  
XX  
XX AC ABV24425;  
XX  
XX 16-SEP-2002 (first entry)  
XX  
XX DE Human prostate expression marker cDNA 24416.  
XX  
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
XX KW pharmacogenomic marker; gene; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX WO200160860-A2.  
XX  
XX 23-AUG-2001.  
XX  
XX 20-FEB-2001; 2001WO-US005171.  
XX  
XX 17-FEB-2000; 2000US-0183319P.  
XX PR 16-MAR-2000; 2000US-0189862P.  
XX PR 25-MAY-2000; 2000US-0207454P.  
XX PR 09-JUN-2000; 2000US-0211314P.  
XX PR 18-JUL-2000; 2000US-0219007P.  
XX PR 13-DEC-2000; 2000US-0255281P.  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
XX Schlegel R, Endege WO, Monahan JE;  
XX  
XX WPI; 2001-662795/76.  
XX  
XX PT Novel isolated nucleic acid molecule associated with cancerous state of  
XX prostate cells and correlating with presence of prostate cancer; useful  
XX for detecting presence of prostate cancer, stage of prostate cancer.  
XX  
XX PS Claim 1; Page 4603; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
XX specification or its complement. (I) is useful for: (a) assessing whether  
XX a patient is afflicted with prostate cancer; (b) monitoring the  
XX progression of prostate cancer in a patient; (c) assessing the efficacy  
XX of a test compound to inhibit prostate cancer in a patient; (d) assessing  
XX the efficacy of a therapy for inhibiting prostate cancer in a patient;  
XX (e) selecting a composition for inhibiting prostate cancer in a patient;  
XX (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
XX determining whether prostate cancer has metastasized in a patient; (h)  
XX assessing the aggressiveness or indolence of prostate cancer in a patient  
XX ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX  
SQ Sequence 2407 BP; 781 A; 402 C; 526 G; 694 T; 0 U; 4 Other;  
  
Query Match 22.8%; Score 426; DB 5; Length 2407;  
Best Local Similarity 100.0%; Pred. No. 2.4e-158;  
Matches 426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 582 AGACCAATCTGACGCTGGAATTAACCAATGACAGTCTCTTACTTCACTTCAAGAAAGTC 641  
DB 1671 AGACCAATCTGACGCTGGAATTAACCAATGACAGTCTCTTACTTCACTTCAAGAAAGTC 1730  
  
QY 642 TGAAGGCGGAAAGAAAGCAAGAAAGGAGCACTTCAAGGACCAAGGAGGTGGAAGT 701  
DB 1731 TGAAGGCGGAAAGAAAGCAAGAAAGGAGCACTTCAAGGACCAAGGAGGTGGAAGT 1790  
  
QY 702 GCTGAGTATGAATTAATCTGATGAATATTTGACTCTGCTGTAAGAAAGTAAAGATGTT 761  
DB 1791 GCTGAGTATGAATTAATCTGATGAATATTTGACTCTGCTGTAAGAAAGTAAAGATGTT 1850  
  
QY 762 ATCGAAATCTACATGGAATTAATGGAAGTCCCTTCAAGTTTGAAGTAAAGCATTTTGAAGC 821  
DB 1851 ATCGAAATCTACATGGAATTAATGGAAGTCCCTTCAAGTTTGAAGTAAAGCATTTTGAAGC 1910

OY	822	AAATTAAGGAATTCACCTTTGTAAGTGTGAACTAATCCCTAATATGATAGGTTT	881
Db	1911	AAATTAAGGAATTCACCTTTGTAAGTGTGAACTAATCCCTAATATGATAGGTTT	1970
OY	882	ATATTGATTTCATGGGTACAGGTCATATAATTTTGGAACTAGATGCTGAATAT	941
Db	1971	ATATTGATTTCATGGGTACAGGTCATATAATTTTGGAACTAGATGCTGAATAT	2030
OY	942	CAAGGAGACAGCCATAGTCTCTTACAGTGCCCTGTGGTGTCTCTCAAACTGAATTGG	1001
Db	2031	CAAGGAGACAGCCATAGTCTCTTACAGTGCCCTGTGGTGTCTCTCAAACTGAATTGG	2090
OY	1002	GTGGGA	1007
Db	2091	GTGGGA	2096

Search completed: July 29, 2004, 05:54:36  
 Job time : 767 secs



This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2004, 05:04:37 ; Search time 4998 Seconds

(without alignments)  
11184.874 Million cell updates/sec

Title: US-09-787-491B-18

Perfect score: 1872

Sequence: 1 gccctctagcgtgtgtgtc.....tatcaaaaaaaaaaaaaa 1872

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estma:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_esthum:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_man:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrt:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	794	42.4	914	13	BX352678 BX352678
2	776	41.5	867	13	BX392494 BX392494
3	769	41.1	1106	13	BX363934 BX363934
4	754	40.3	1201	13	BX356666 BX356666

5	722	38.6	858	9	AU130178 AU130178
6	699	37.3	866	13	BX326041 BX326041
7	690	36.9	926	13	BX450740 BX450740
8	677	36.2	879	13	BQ421218 AGENCOURT
9	657	35.1	1035	12	BM928059 AGENCOURT
10	652	34.8	788	12	BG611131 AGENCOURT
11	642	34.3	710	13	BM683427 UI-CF-RCL
12	640	34.2	1038	12	BM921276 UI-CF-RCL
13	637	34.0	895	13	BX408797 UI-CF-RCL
14	603	32.2	670	13	BU621746 UI-E-FOI-
15	596	31.8	760	12	BM728917 UI-E-FOI-
16	595	31.8	761	12	BI836024 UI-E-FOI-
17	580	31.0	679	13	BU742938 UI-E-FOI-
18	573	30.6	727	14	CB053447 NISC_g114
19	560	29.9	622	12	BM836931 K-EST0112
20	560	29.9	871	14	CD512033 AGENCOURT
21	560	29.9	1095	12	BM910977 AGENCOURT
22	557	29.8	786	10	BE281165 601158351
23	556	29.7	912	13	BQ220979 AGENCOURT
24	553	29.5	579	13	BU689661 UI-CF-RNO
25	551	29.4	1229	12	BM467160 AGENCOURT
26	548	29.3	686	13	BX509477 DKP2P686h
27	539	28.8	673	14	CD690132 EST6655_h
28	539	28.8	722	13	BU928568 AGENCOURT
29	534	28.5	823	12	BI559513 603252842
30	530	28.3	544	12	BM147285 TCAP1010
31	530	28.3	648	12	BM837276 K-EST0113
32	530	28.3	901	13	BU193345 AGENCOURT
33	530	28.3	1114	12	BM544074 AGENCOURT
34	529	28.3	536	10	BE018162 b076a11.y
35	529	28.3	1068	12	BM456685 AGENCOURT
36	527	28.2	527	14	CB112315 K-EST0154
37	521	27.8	728	10	BE259920 601184810
38	515	27.5	917	13	BX388512 BX388512
39	511	27.3	701	10	BE295413 601175939
40	511	27.3	767	10	BF691855 602248438
41	510	27.2	928	13	BQ428125 AGENCOURT
42	509	27.2	702	9	AU132812 AU132812
43	507	27.1	513	10	BE301644 b076a11.x
44	503	26.9	810	13	BU599405 AGENCOURT
45	499	26.7	699	12	BG532282 602561373

#### ALIGNMENTS

RESULT 1  
LOCUS BX352678 914 bp mRNA linear EST 02-MAY-2003  
DEFINITION BX352678 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens  
CDNA clone CS0DB001YG15 5-PRIME, mRNA sequence.  
ACCESSION BX352678  
VERSION BX352678.1 GI:30347694  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 914)  
AUTHORS Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: segre@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 6566.r  
Contact: Feng Liang Email: fliang@life.com URL:  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID: CS0DB001AD08QPL.

#### FEATURES

source

l. .914



Db 507 CTAGTGAATCATTACAGTATGGCCATCTGTTTGTGGAGTGCTTATTCATACAT 448  
Qy 1412 ATTTACACCACTTCGTATCAATGTATATATAGAACACATATACATCAAGATAATTA 1471  
Db 447 ATTTACACCACTTCGTATCAATGTATATATAGAACACATATACATCAAGATAATTA 388  
Qy 1472 TTGTGTGGTATCTGCCATTTTAAAGTATCCGATTTTGTATCAATATATATATATAT 1531  
Db 387 TTGTGTGGTATCTGCCATTTTAAAGTATCCGATTTTGTATCAATATATATATATAT 328  
Qy 1532 AAAAATGATTTATCTGTATATAACTGGTTATTTGTGAGAGCACTGTATATCTTAAG 1591  
Db 327 AAAAATGATTTATCTGTATATAACTGGTTATTTGTGAGAGCACTGTATATCTTAAG 268  
Qy 1592 TTATATATATTTTACTCTGCTCCACCAAAACATGCTAGATATATACCCCAAAATA 1651  
Db 267 TTATATATATTTTACTCTGCTCCACCAAAACATGCTAGATATATACCCCAAAATA 208  
Qy 1652 AGTATTTTAACTTTGATTAAGTATTAAGAGACCTGGTGCTATATATATATATATTTGAG 1711  
Db 207 AGTATTTTAACTTTGATTAAGTATTAAGAGACCTGGTGCTATATATATATATATTTGAG 148  
Qy 1712 GCAGACAGAGAGCTGTATACCTAACTGATTTAGTATGTATTTAGATTAAGAAAATGTTCA 1771  
Db 147 GCAGACAGAGAGCTGTATACCTAACTGATTTAGTATGTATTTAGATTAAGAAAATGTTCA 88  
Qy 1772 CCAATTTTACTTTTATGATTTATCATGTATCATTTTATAGGGAGCA 1818  
Db 87 CCAATTTTACTTTTATGATTTATCATGTATCATTTTATAGGGAGCA 41

RESULT 3  
BX363934 1106 bp mRNA linear EST 05-MAY-2003  
LOCUS BX363934 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
DEFINITION Homo sapiens cDNA clone CS0DL007YC18 5-PRIME, mRNA sequence.  
ACCESSION BX363934  
VERSION BX363934.1 GI:30378776  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1106)  
AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

FEATURES  
Source  
1..1106  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DL007YC18"  
/cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/cell\_line="RAMOS CELL LINE"  
/clone\_id="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN  
Query Match 41.1%; Score 769; Db 13; Length 1106;

Best Local Similarity 99.8%; Pred. No. 3,1e-170;  
Matches 939; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
Qy 70 GCCCGCCGAGACACGCGCCGCCATGTCCTCCGCTCCCGCCCAACGTCCTG 129  
Db 87 GCCCGCCGAGACACGCGCCGCCATGTCCTCCGCTCCCGCCCAACGTCCTG 146  
Qy 130 TTGCTCAGGAAAGTGGCCGACGACACGAGTCTGAAGCTTGGGCGGAAATTTGGT 189  
Db 147 TTGCTCAGGAAAGTGGCCGACGACACGAGTCTGAAGCTTGGGCGGAAATTTGGT 206  
Qy 190 TATGCTCCTATAGTATGATGTATGTTCCACTTGAATTTCTACACTGCGCTCAAGAGA 249  
Db 207 TATGCTCCTATAGTATGATGTATGTTCCACTTGAATTTCTACACTGCGCTCAAGAGA 266  
Qy 250 TTGCTTATGTTCAATTTGAGAGTTCGTATGCTGAAGCGCTTTACATATTTGGAC 309  
Db 267 TTGCTTATGTTCAATTTGAGAGTTCGTATGCTGAAGCGCTTTACATATTTGGAC 326  
Qy 310 AGAAAGTGAAATTTGAGACGGCAATGAATACAGTTTGCCGAGGGGATCGAAGACA 369  
Db 327 AGAAAGTGAAATTTGAGACGGCAATGAATACAGTTTGCCGAGGGGATCGAAGACA 386  
Qy 370 CCAATATGATGAAAGCCCAAGAAAGGAGAGATGTATACAGTTCTTCAAGCTATGAT 429  
Db 387 CCAATATGATGAAAGCCCAAGAAAGGAGAGATGTATACAGTTCTTCAAGCTATGAT 446  
Qy 430 TATGACAGATCAGACGTTCTTGAAGCCGAAAGTTATGAAGAAGAGATCAAGAGTGG 489  
Db 447 TATGACAGATCAGACGTTCTTGAAGCCGAAAGTTATGAAGAAGAGATCAAGAGTGG 506  
Qy 490 TCTTTGATTTACACTATGAAAGATCGTATGCTCTGAAACAGTATGACCGACCTGAGAGA 549  
Db 507 TCTTTGATTTACACTATGAAAGATCGTATGCTCTGAAACAGTATGACCGACCTGAGAGA 566  
Qy 550 CCACGGCGTATAGA -AGCCATTCGACATGATGACCCAACTGAGATATCCAG 608  
Db 567 CCACGGCGTATAGAAGGCGATTCGACATGATGACCCAACTGAGATATCCAG 626  
Qy 609 TACAGTTCTGCTTCTACACTTCAAGAAAGATCTGAAAGCGGAAAGAACCCAAAGAG 668  
Db 627 TACAGTTCTGCTTCTACACTTCAAGAAAGATCTGAAAGCGGAAAGAACCCAAAGAG 686  
Qy 669 GCAAGTCAAGGACCAAGAGGTGGGTGAAGGTCTGAGATGATATCTAGGAATAT 728  
Db 687 GCAAGTCAAGGACCAAGAGGTGGGTGAAGGTCTGAGATGATATCTAGGAATAT 746  
Qy 729 TTGACTCTGCTGAAAGAGTAAAGATGTTATCGAAAATCATGATATATTTGAAG 788  
Db 747 TTGACTCTGCTGAAAGAGTAAAGATGTTATCGAAAATCATGATATATTTGAAG 806  
Qy 789 TCCCTTCAAGTTTGAAGTAAAGCTTTTGAAGACAAATTAAGAAATTTCAACTTTGACT 848  
Db 807 TCCCTTCAAGTTTGAAGTAAAGCTTTTGAAGACAAATTAAGAAATTTCAACTTTGACT 866  
Qy 849 TGTGAAACTATATCCCTAAATATGATAGGTTTATATGATTCATGGGTAAACAGGTGAT 908  
Db 867 TGTGAAACTATATCCCTAAATATGATAGGTTTATATGATTCATGGGTAAACAGGTGAT 926  
Qy 909 AATTAATTAATTTGAAACTAGAGATGTCTGAATATCAAGAAAGACGCAATGCTCTTACA 968  
Db 927 AATTAATTAATTTGAAACTAGAGATGTCTGAATATCAAGAAAGACGCAATGCTCTTACA 986  
Qy 969 GTGCTCTGTGCTGCTCTCAACATCAATTTGGGTGGGAAA 1009  
Db 987 GTGCTCTGTGCTGCTCTCAACATCAATTTGGGTGGGAAA 1027

RESULT 4  
BX356666/c 1201 bp mRNA linear EST 05-MAY-2003  
LOCUS BX356666 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
DEFINITION Clone CS0DI015D03 3-PRIME, mRNA sequence.

ACCESSION BX356666  
 VERSION BX356666.1 GI:30370049  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6566.x  
 Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID: CS0D1015CB02NP1.  
 FEATURES  
 source  
 1..1201  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1015YD03"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 40.3%; Score 754; DB 13; Length 1201;  
 Best Local Similarity 99.7%; Pred. No. 9e-167;  
 Matches 974; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

853 GAACTAATCCCTAATATGAATAGTTTATATGATTGATGGGTACAGGTCCTAATA 912  
 |||||  
 993 GAACTAATCCCTAATATGAATAGTTTATGATTCATGGGTACAGGTCCTAATA 934  
 |||||  
 913 AATTATGGAACCTGGATGTCGAATATCAAGGAAGACGATGCTCTTACAGTGC 972  
 |||||  
 933 AATTATGGAACCTGGATGTCGAATATCAAGGAAGACGATGCTCTTACAGTGC 874  
 |||||  
 973 CTCTGTGCTGCTCAACTGAATGGGTGGGAAAGATAGTCCATATATAAAGTT 1032  
 |||||  
 873 CTCTGTGCTGCTCAACTGAATGGGTGGGAAAGATAGTCCATATATAAAGTT 814  
 |||||  
 1033 CCATTTTGGCATTAATGGCAAACTCTGGCTTTGTTATTTGGTCCAGTCTTCTGC 1092  
 |||||  
 813 CCATTTTGGCATTAATGGCAAACTCTGGCTTTGTTATTTGGTCCAGTCTTCTGC 754  
 |||||  
 1093 TTAATCATTTGCTTTGTGGACATCTGTGTTATTTACTGTACACACAGCAGTATTAA 1152  
 |||||  
 753 TTAATCATTTGCTTTGTGGACATCTGTGTTATTTACTGTACACACAGCAGTATTAA 694  
 |||||  
 1153 TCTGTCTTAATACTCTCTTCCAGTAATTCGAATTAATTTGACATCCAGCTAAGAG 1212  
 |||||  
 693 TCTGTCTTAATACTCTCTTCCAGTAATTCGAATTAATTTGACATCCAGCTAAGAG 634  
 |||||  
 1213 GCCCATCTCTTGCACACTCTTCTTCTAGTCAGATATTTGACGAAATTTTATGAGCCCT 1272  
 |||||  
 633 GCCCATCTCTTGCACACTCTTCTTCTAGTCAGATATTTGACGAAATTTTATGAGCCCT 574  
 |||||  
 1273 ACTGTGGGCAAAATCATTTGACTAGGATATTTGAGAAATATAGATATTCCTTATTCAGTA 1332  
 |||||  
 573 ACTGTGGGCAAAATCATTTGACTAGGATATTTGAGAAATATAGATATTCCTTATTCAGTA 514  
 |||||  
 1333 AA-TGCTACTAGGACCAATCTAGTGAATCATTTACAGTATGGCTCATTTGTTTGA 1391  
 |||||  
 513 AATTGTTTTCGTAGGACCAATCTAGTGAATCATTTACAGTATGGCTCATTTGTTTGA 454  
 |||||

QY 1392 GGTGTTTATTCATACATATTTTACCACTGCTATCATATGTAATTATAGACAAAT 1451  
 Db 453 GGTGTTTATTCATACAAATATTTTACCACTGCTATCATATGTAATTATAGACAAAT 394  
 QY 1452 ATAGATCAAGGATTAAGTAATTTGTTGTTATCTGCCATTTTAAAGTATCCAGTATTGA 1511  
 Db 393 ATAGATCAAGGATTAAGTAATTTGTTGTTATCTGCCATTTTAAAGTATCCAGTATTGA 334  
 QY 1512 TCACATTTATTAATTAATGAAAAATGATTTAATCTGTAATTAACCTGGTTATTGTGCA 1571  
 Db 333 TCACATTTATTAATTAATGAAAAATGATTTAATCTGTAATTAACCTGGTTATTGTGCA 274  
 QY 1572 GTGACTGTAATTAATTAAGTATTAATTAATTTGTTATCTGCCATTTTAAAGGACATGCT 1631  
 Db 273 GTGACTGTAATTAATTAAGTATTAATTAATTTGTTATCTGCCATTTTAAAGGACATGCT 214  
 QY 1632 AGGATATTAACCCCAAAATTAATTAATTTAATCTTTGATTAAGTATTAAGGACATGCTGTC 1691  
 Db 213 AGGATATTAACCCCAAAATTAATTAATTTAATCTTTGATTAAGTATTAAGGACATGCTGTC 154  
 QY 1692 TATTAATTAATTTTATTTGAGGACAGACAGAGCTGTATCTTAATGATTTAGTATGTTTC 1751  
 Db 153 TATTAATTAATTTTATTTGAGGACAGACAGAGCTGTATCTTAATGATTTAGTATGTTTC 94  
 QY 1752 TGTATTTGAGAAATGTTCAACCAATTAATTAATCTTTTATGATTTACATGATTAATTTATA 1811  
 Db 93 TGTATTTGAGAAATGTTCAACCAATTAATTAATCTTTTATGATTTACATGATTAATTTATA 34  
 QY 1812 GGGGACATGTTCTGTGT 1828  
 Db 33 GGGGACATGTTCTGTGT 17

RESULT 5  
 AUI30178 858 bp mRNA linear EST 01-AUG-2002  
 LOCUS AUI30178 NT2RP3 Homo sapiens cDNA clone NT2RP3000370 5', mRNA  
 DEFINITION  
 sequence.  
 ACCESSION AUI30178  
 VERSION AUI30178.1 GI:10990532  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 858)  
 AUTHORS Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, T., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and Iwagai, T.  
 TITLE HRI human cDNA project  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Takao Iwagai  
 Genomics Laboratory  
 Helix Research Institute  
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 Tel: 81-438-52-3975  
 Fax: 81-438-52-3986  
 Email: genomics@hri.co.jp  
 HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.  
 FEATURES  
 source  
 1..858  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="NT2RP3000370"  
 /cell\_type="teratocarcinoma"  
 /cell\_line="NT2"  
 /clone\_lib="NT2RP3"  
 /note="Vector: PME18SFL3; mRNA from NT2 neuronal precursor"

ORIGIN cells after 2-weeks retinoic acid (RA) induction"

Query Match 38.6%; Score 722; DB 9; Length 858;  
Best Local Similarity 100.0%; Pred. No. 3,4e-159; Indels 0; Gaps 0;  
Matches 722; Conservative 0; Mismatches 0;

QY 1080 CAGTGTTCCTGCTTAATCATTTGCTTGTGGCATCTGTGTTATTTACTGTACACCA 1139  
DB 1 CAGTGTTCCTGCTTAATCATTTGCTTGTGGCATCTGTGTTATTTACTGTACACCA 60  
QY 1140 CAGGAGTTTACATCTGTCTTAACCTCTCCCGAGTAAATTCGAATTAATTTGACA 1199  
DB 61 CAGGAGTTTACATCTGTCTTAACCTCTCCCGAGTAAATTCGAATTAATTTGACA 120  
QY 1200 TCCAGCTAAGAGGAGCCCATCTCTTCCACCTCTTCCCTAGTATATTGACAAAT 1259  
DB 121 TCCAGCTAAGAGGAGCCCATCTCTTCCACCTCTTCCCTAGTATATTGACAAAT 180  
QY 1260 TTATTGAGCCCTTACTGTGGGCAAAATCATTTGATGATTAATGAAAAATGATTAAT 1319  
DB 181 TTATTGAGCCCTTACTGTGGGCAAAATCATTTGATGATTAATGAAAAATGATTAAT 240  
QY 1320 CCTTATTCAGTAATATGCTACTGAGCACAATCTGATATCATTAACATGAGCTCAT 1379  
DB 241 CCTTATTCAGTAATATGCTACTGAGCACAATCTGATGATTAATGAGTATGAGCTCAT 300  
QY 1360 TGTGTTGTTGAGTGTGTTATTCATTAACAATATTTTACACCATTCGATCAATGTAAT 1439  
DB 301 TGTGTTGTTGAGTGTGTTATTCATTAACAATATTTTACACCATTCGATCAATGTAAT 360  
QY 1440 ATGAGCAACAATATGATCAAGATTAAGTATGTTGTTATTCGATTAATTAAGTA 1499  
DB 361 ATGAGCAACAATATGATCAAGATTAAGTATGTTGTTATTCGATTAATTAAGTA 420  
QY 1500 TCCAGATTTGATGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1559  
DB 421 TCCAGATTTGATGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 480  
QY 1560 GTTTATTTGAGTGTGTTATTCATTAACAATATTTTACACCATTCGATCAATGTAAT 1619  
DB 481 GTTTATTTGAGTGTGTTATTCATTAACAATATTTTACACCATTCGATCAATGTAAT 540  
QY 1620 CAACACATGCTAGGATTAACCCGCCAAATTAAGTATTAATTTGATTAATTAAG 1679  
DB 541 CAACACATGCTAGGATTAACCCGCCAAATTAAGTATTAATTTGATTAATTAAG 600  
QY 1680 GAGAGTGGGTGCTAATTAATTAATTTTGAAGCAGACAGAGCTGTTATCCAACTGA 1739  
DB 601 GAGAGTGGGTGCTAATTAATTAATTTTGAAGCAGACAGAGCTGTTATCCAACTGA 660  
QY 1740 TTATGATGTTCTGTAATTTGAAGAAATGTCCACAAATTAATTTTATGATTAAT 1799  
DB 661 TTATGATGTTCTGTAATTTGAAGAAATGTCCACAAATTAATTTTATGATTAAT 720  
QY 1800 GT 1801  
DB 721 GT 722

RESULT 6  
EX326041/c 866 bp mRNA linear EST 02-MAY-2003  
LOCUS EX326041 Homo sapiens NEUROBLASTOMA COR 10-NORMALIZED Homo sapiens  
DEFINITION CDNA clone CS0DB001YGL5 3-PRIME, mRNA sequence.  
ACCESSION EX326041  
VERSION BX326041.1 GI:30344506  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 866)

# AUTHORS

Li, M.B., Gruber, C., Jesse, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)

## JOURNAL

Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France

## COMMENT

Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by life technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 6566.r  
Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0AU002ZC11\_U0179\_1.  
Location/Qualifiers  
1..866  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DB001YGL5"  
/tissue\_type="NEUROBLASTOMA COR 10-NORMALIZED"  
/clone\_id="Homo sapiens NEUROBLASTOMA COR 10-NORMALIZED"  
/note="First strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

## FEATURES

source

## ORIGIN

Query Match 37.3%; Score 699; DB 13; Length 866;  
Best Local Similarity 99.9%; Pred. No. 7.6e-154; Indels 0; Gaps 0;  
Matches 749; Conservative 0; Mismatches 1;

QY 561 AGAAGCCATTCGCAATGATGACCAACATGCACTGGAATACCGATGACGTTCTGCT 620  
DB 788 AGAAGCCATTCGCAATGATGACCAACATGCACTGGAATACCGATGACGTTCTGCT 729  
QY 621 TACTACCTTCAAGAAAGATCTGAAAGCGGAAAAAGAAACCAAGAGGCGATTCAGCG 680  
DB 728 TACTACCTTCAAGAAAGATCTGAAAGCGGAAAAAGAAACCAAGAGGCGATTCAGCG 669  
QY 681 ACCAAGGGTGGGAGGAGGCTGCGCATGATGATCTGTACGATATTTTGACTCTGCT 740  
DB 668 ACCAAGGGTGGGAGGAGGCTGCGCATGATGATCTGTACGATATTTTGACTCTGCT 609  
QY 741 CTGAAAGATTAAGAAATGTTATGCAAAACATGATGAATTAATTAATTAATTAATTAAT 800  
DB 608 CTGAAAGATTAAGAAATGTTATGCAAAACATGATGAATTAATTAATTAATTAATTAAT 549  
QY 801 TGAAGTATGATTTTGAAGCAAAATTAAGAAATTAATTAATTAATTAATTAATTAAT 860  
DB 548 TGAAGTATGATTTTGAAGCAAAATTAAGAAATTAATTAATTAATTAATTAATTAAT 489  
QY 861 TCCCTAATATGAATTAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 920  
DB 488 TCCCTAATATGAATTAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 429  
QY 921 GAACTAGATGCTGGAATATCAAGAAAGACAGCATAGTCTCTTACAGTCCCTGTTG 980  
DB 428 GAACTAGATGCTGGAATATCAAGAAAGACAGCATAGTCTCTTACAGTCCCTGTTG 369  
QY 981 GTCTGTCTCAAACTGAATTTGGGTGGGAAAGGTAATGTCGAATTAATTAATTAATTTT 1040  
DB 368 GTCTGTCTCAAACTGAATTTGGGTGGGAAAGGTAATGTCGAATTAATTAATTAATTTT 309  
QY 1041 GCAATTAATGGAATCTGCTGTTTATTTTATTTTGGGCGAGTCTTCTGCTTAATCAT 1100  
DB 308 GCAATTAATGGAATCTGCTGTTTATTTTATTTTGGGCGAGTCTTCTGCTTAATCAT 249  
QY 1101 TTGCTTTGTTGGCATCTGTGTTTATTTACTTTGACACCAATGAGATTAATCATCTGCTT 1160  
DB 248 TTGCTTTGTTGGCATCTGTGTTTATTTACTTTGACACCAATGAGATTAATCATCTGCTT 189  
QY 1161 AACTACTCTTCCAGGTAATTCGAATTAATTTGATTCATTCAGTAAAGAGGCCCATCT 1220  
DB 188 AACTACTCTTCCAGGTAATTCGAATTAATTTGATTCATTCAGTAAAGAGGCCCATCT 129

QY 1221 CTTCTCACTCTTCTTCTAGTCGTATATTCAGCAATATTTATGAGCCCTTACTGTGGG 1280  
 DB 128 CTTCTCACTCTTCTTCTAGTCGTATATTCAGCAATATTTATGAGCCCTTACTGTGGG 69  
 QY 1281 CAAATCATGTACTGATATGAGAAAA 1310  
 DB 68 CAAATCATGTACTGATATGAGAAAA 39

RESULT 7  
 BX450740 926 bp mRNA linear EST 22-MAY-2003  
 LOCUS BX450740 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA  
 DEFINITION clone CS0D5001YH09 5-PRIME, mRNA sequence.  
 ACCESSION BX450740 GI:31022152  
 VERSION BX450740.1  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 926)  
 AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: sequef@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 798.r for  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0BA011ZH09\_CS01033\_1cluster=798.r.  
 Contact: Feng Liang Email: fliang@life.techn.com URL:  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID: CS0BA011ZH09\_CS01033\_1.

FEATURES  
 source  
 Location/Qualifiers  
 1..926  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D5001YH09"  
 /tissue\_type="B CELLS (RAMOS CELL LINE)"  
 /cell\_line="RAMOS CELL LINE"  
 /clone\_1lb="Homo sapiens B CELLS (RAMOS CELL LINE)"  
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN  
 Query Match 36.9%; Score 690; DB 13; Length 926;  
 Best Local Similarity 100.0%; Pred. No. 8, 9e-152;  
 Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 AATAATGAGTCCCTTCAAGTTTGAAGTAAAGTATTTAGACAAATAAAGAAATTC 838  
 DB 21 AATAATGAGTCCCTTCAAGTTTGAAGTAAAGTATTTAGACAAATAAAGAAATTC 80  
 QY 839 ACTTTGACTGTGGAACTTAATCCCTAAATATGATAGCTTATATTTGATTCATGGGTA 898  
 DB 81 ACTTTGACTGTGGAACTTAATCCCTAAATATGATAGCTTATATTTGATTCATGGGTA 140  
 QY 899 ACAGGTCCATATAAATTTATTTGAAACTAGATGTCTGAATATCAAGAGACAGCCATA 958  
 DB 141 ACAGGTCCATATAAATTTATTTGAAACTAGATGTCTGAATATCAAGAGACAGCCATA 200  
 QY 959 GTCTCTTACAGTCTCTGTGTGTCTGTCTCAATGCAATGGTGGGAAAAAGGTATGT 1018  
 DB 201 GTCTCTTACAGTCTCTGTGTGTGTCTGTCTCAATGCAATGGTGGGAAAAAGGTATGT 260

QY 1019 CCAATATAAAGTCCATTTTGGCATATTTGGCAAAATCTGCTTTATTTTGGTG 1078  
 DB 261 CCAATATAAAGTCCATTTTGGCATATTTGGCAAAATCTGCTTTATTTTGGTG 320  
 QY 1079 CCAAGTCTTCTGCTTAAATCATATTTGCTTTGTTGGCATCTGTGTTATTTACTTGAAC 1138  
 DB 321 CCAAGTCTTCTGCTTAAATCATATTTGCTTTGTTGGCATCTGTGTTATTTACTTGAAC 380  
 QY 1139 ACATGCAAGTTTACATCTGTCTTAACTACTCTCTTCCAGGTAATTCATTTATTTGAC 1198  
 DB 381 ACATGCAAGTTTACATCTGTCTTAACTACTCTCTTCCAGGTAATTCATTTATTTGAC 440  
 QY 1199 ATCCAGCTTAAGAGGCGCCATCTCTCTGACCTCTTCTCCAGTCAAGTAATTCAGCAATA 1258  
 DB 441 ATCCAGCTTAAGAGGCGCCATCTCTCTGACCTCTTCTCCAGTCAAGTAATTCAGCAATA 500  
 QY 1259 TTTATTTAGACCTTACTCTGTGGCAATCATTTGTACTGTATTTGAAGAAAAATGATTAAT 1318  
 DB 501 TTTATTTAGACCTTACTCTGTGGCAATCATTTGTACTGTATTTGAAGAAAAATGATTAAT 560  
 QY 1319 TCCCTTATTCAGTAAATGCTTCTCTGAGCAGCATCTTATGTAATCATTTAGCTATGCGCTCA 1378  
 DB 561 TCCCTTATTCAGTAAATGCTTCTCTGAGCAGCATCTTATGTAATCATTTAGCTATGCGCTCA 620  
 QY 1379 TTGTTTGTGTTGAGGTGTGTATTTCAATAAATTTTACACATTCGATCAATGTAAAT 1438  
 DB 621 TTGTTTGTGTTGAGGTGTGTATTTCAATAAATTTTACACATTCGATCAATGTAAAT 680  
 QY 1439 TATAGAACAAATATATCATCATCAAGATTAAG 1468  
 DB 681 TATAGAACAAATATATCATCATCAAGATTAAG 710

RESULT 8  
 BQ421218 879 bp mRNA linear EST 23-MAY-2002  
 LOCUS BQ421218  
 DEFINITION AGENCOURT 7763286 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:6013472  
 5', mRNA sequence.  
 ACCESSION BQ421218  
 KEYWORD BQ421218.1 GI:21116533  
 VERSION EST.  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 879)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE NIH-MGC National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LINL)  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:  
 http://image.llnl.gov  
 Plate: LLAM13206 row: g column: 09  
 High quality sequence stop: 679.  
 Location/Qualifiers  
 1..879  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6013472"  
 /tissue\_type="embryonal carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_1lb="NIH MGC 92"  
 /note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; Oligo-dT primed. Average insert size 2.5 kb. Library enriched for

## ORIGIN

full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC library."

Query Match 36.2%; Score 677; DB 13; Length 879;  
Best Local Similarity 99.7%; Pred. No. 9.9e-149;  
Matches 777; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 932 GTCGATATCAAGGAAGACAGCATAGTCTTACAGTGCCTGTGTGTCTCAA 991
DB 5 GTCGATATCAAGGAAGACAGCATAGTCTTACAGTGCCTGTGTGTCTCAA 64
QY 992 ACTGAATGGGTGGGAAAAGCATAGTCCATATTAAGTTCATTTTGGCATTAATG 1051
DB 65 ACTGAATGGGTGGGAAAAGCATAGTCCATATTAAGTTCATTTTGGCATTAATG 124
QY 1052 CAAATCTTGCCCTTTGTTATTTTGGTGGCCAGTGTTCCTGCTTAACATTTGTTG 1111
DB 125 CAAATCTTGCCCTTTGTTATTTTGGTGGCCAGTGTTCCTGCTTAACATTTGTTG 184
QY 1112 GCATCTGTGTTATTTACTTGTACACAGCATGATTTTACATCTGTCTTACACTCTT 1171
DB 185 GCATCTGTGTTATTTACTTGTACACAGCATGATTTTACATCTGTCTTACACTCTT 244
QY 1172 CCCAGGTAATTCOAATTTATTTGATACACCGCTAAGGGGCCCATCTCTCTCACTC 1231
DB 245 CCCAGGTAATTCOAATTTATTTGATACACCGCTAAGGGGCCCATCTCTCTCACTC 304
QY 1232 TTTCCTAGTCAATATTCAGCAAAATATTTATTTGAGCCCTTACTGTGGGCAATATG 1291
DB 305 TTTCCTAGTCAATATTCAGCAAAATATTTATTTGAGCCCTTACTGTGGGCAATATG 364
QY 1292 ACTGATTAATTGAGAAAATAGATTAATTCCTTATTCAGTAATGTCTACTGAGCAAT 1351
DB 365 ACTGATTAATTGAGAAAATAGATTAATTCCTTATTCAGTAATGTCTACTGAGCAAT 424
QY 1352 CTAGGATCAATTAAGTATGAGCCCATTTGTTGTTGAGGTGTTATTCATAACAT 1411
DB 425 CTAGGATCAATTAAGTATGAGCCCATTTGTTGTTGAGGTGTTATTCATAACAT 484
QY 1412 ATTTTACACATTCCTATCAATGTATTTATAGAACACAAATATAGATCAAGGATAGTA 1471
DB 485 ATTTTACACATTCCTATCAATGTATTTATAGAACACAAATATAGATCAAGGATAGTA 544
QY 1472 TTGTGTGTATTCGCAATTTAAAGTATTCAGATTTTATTCATTAATTAATATG 1531
DB 545 TTGTGTGTATTCGCAATTTAAAGTATTCAGATTTTATTCATTAATTAATATG 604
QY 1532 AAAAATGATTTATTCGTATTAATTAAGTATTTATTCAGTGTATTAATTAATAG 1591
DB 605 AAAAATGATTTATTCGTATTAATTAAGTATTTATTCAGTGTATTAATTAATAG 664
QY 1592 TTATATTAATTTGTTTACTCTGCTCACCAACACATCTAGAGATTAATTAATTA 1651
DB 665 TTATATTAATTTGTTTACTCTGCTCACCAACACATCTAGAGATTAATTAATTA 724
QY 1652 AGTATTTTACTTTGATTAAGTATTAAGAGACTGGGTCTAATTAATTAATTTGA 1710
DB 725 GGTATTTTAACTTTCATTAAGTATTAAGAGACTGGGTCTAATTAATTAATTTGA 783
```

RESULT 9 1035 bp mRNA linear EST 12-MAR-2002  
BM928059  
LOCUS BM928059  
DEFINITION AGENCOURT 6728747 NIH\_MGC\_100 Homo sapiens cDNA IMAGE:5797875  
5', mRNA sequence.  
ACCESSION BM928059  
VERSION BM928059.1 GI:19378438  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1035)  
AUTHORS NIH-MGC  
TITLE NIH-MGC htp://img.ncbi.nih.gov/  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: CGAP (Stanford)  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
http://image.llnl.gov  
Plate: LICM2024 row: p column: 04  
High quality sequence stop: 678.  
Location/Qualifiers  
1. 1035  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5797875"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_100"  
/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-CDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

## FEATURES

## source

## ORIGIN

Query Match 35.1%; Score 657; DB 12; Length 1035;  
Best Local Similarity 99.9%; Pred. No. 3.9e-144;  
Matches 777; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
QY 215 TTCCACTGATTTTCTACACTGCGGCTCCAAAGATTTGTTATGTTCAATTTGAGATG 274
DB 1 TTCCACTGATTTTCTACACTGCGGCTCCAAAGATTTGTTATGTTCAATTTGAGATG 60
QY 275 TTCTGATGCTGAAGAGCGCTTTACATAATTGGAAGAAATGATTTGTGAGCGGCGA 334
DB 61 TTCTGATGCTGAAGAGCGCTTTACATAATTGGAAGAAATGATTTGTGAGCGGCGA 120
QY 335 TTGAATATCAGTTTCCCGAGGGGATCGAAAGACCAAAATCAATGAAAGCCAAAGAG 394
DB 121 TTGAATATCAGTTTCCCGAGGGGATCGAAAGACCAAAATCAATGAAAGCCAAAGAG 180
QY 395 GGAGGAATGTGTACAGTTTCTTACGCTATGTGATTTATGACAGATATACAGCTTCTAGA 454
DB 181 GGAGGAATGTGTACAGTTTCTTACGCTATGTGATTTATGACAGATATACAGCTTCTAGA 240
QY 455 GCCAAGTTATGAAGGAGGAGATCAAGAAGTGGCTTTTGTATTAACAATTAAGAGAT 514
DB 241 GCCAAGTTATGAAGGAGGAGATCAAGAAGTGGCTTTTGTATTAACAATTAAGAGAT 300
QY 515 GGTATAGTCTTGAAGACATGACCGACTGGAAGACCAAGCGCTAG-AGAAGCCATTCCG 573
DB 301 GGTATAGTCTTGAAGACATGACCGACTGGAAGACCAAGCGCTAGCAAGAACCATTTCCG 360
QY 574 ACATATTAACCAAACTGACAGCTGGAATATCCCACTATCAAGTTCTGTTACTTACACTTCA 633
DB 361 ACATATTAACCAAACTGACAGCTGGAATATCCCACTATCAAGTTCTGTTACTTACACTTCA 420
QY 634 GAAAGATCTGAAGCGGAAAAAGAACCAAGAGGCGAGTTTCAAGCGAACCAAGAGGCGG 693
DB 421 GAAAGATCTGAAGCGGAAAAAGAACCAAGAGGCGAGTTTCAAGCGAACCAAGAGGCGG 480
QY 694 TGGAAGTGTGCAAGTATGATATCTGTACGAATATTTTGACTGTGTTGAAAAAGATAAA 753
```



Db 481 TGAAGTGTGCTGACATGAACTCTGACGAAATTTTGGCTGCTGGAAGATATAA 540  
Qy 754 AGAATGTTATGAAAACATCATGAAATATTTAGTCCCTTCAAGTTTGAAGTAGCAT 813  
Db 541 AGAATGTTATGAAAACATCATGAAATATTTAGTCCCTTCAAGTTTGAAGTAGCAT 600  
Qy 814 TTTAGACAAATTAAGAGAAATTCACCTTGTACTTTGGGAAACCTAATCCCTAAATATGA 873  
Db 601 TTTAGACAAATTAAGAGAAATTCACCTTGTACTTTGGGAAACCTAATCCCTAAATATGA 660  
Qy 874 ATAGGTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 933  
Db 661 ATAGGTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
Qy 934 CTGAATATCAAG 991  
Db 721 CTGAATATCAAG 778

RESULT 10  
Bg611131 788 bp mRNA linear EST 18-APR-2001  
LOCUS 602612089F1 NIH\_MGC\_60 Homo sapiens cDNA clone IMAGE:4737335 5',  
DEFINITION mRNA sequence.  
ACCESSION Bg611131  
VERSION Bg611131.1 GI:13662502  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 788)  
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.  
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DRP  
CDNA Library Preparation: CLONETECH Laboratories, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CMI605 row: b column: 24  
High quality sequence stop: 679.  
Location/Qualifiers  
1. 788  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4737335"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NIH MGC 60"  
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);  
Site\_1: SfiI (ggcggcggcggc); Site\_2: SfiI  
(ggcggcggcggc); Double-stranded cDNA was prepared from  
cell line RNA. 5' and 3' adaptors were used in cloning as  
follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and  
3' adaptor sequence:  
5'-ATTCTAGAGCCGACGAGCGCCGACATG-dt(30)BN-3' (where B = A,  
C, or G and N = A, C, G, or T). Average insert size 1.5  
kb (range 0.9-4.0 kb). 14/15 colonies contained inserts  
by PCR. This library was enriched for full-length clones  
and was constructed by Clontech Laboratories (Palo Alto,  
CA). Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 34.8%; Score 652; DB 12; Length 788;  
Best Local Similarity 99.9%; Pred. No. 7.2e-143;  
Matches 702; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 671 AGTCAAGCCGACCAAGAGGTGGGTGGAAGGTGCTGACATGAAATCTGTAAGATATTT 730  
Db 1 AGTCAAGCCGACCAAGAGGTGGGTGGAAGGTGCTGACATGAAATCTGTAAGATATTT 60  
Qy 731 TGACTGTGGTGTGAAAAGATTAAGATGTTATCGAAAACCTACATGAAATATTTGAAGTC 790  
Db 61 TGACTGTGGTGTGAAAAGATTAAGATGTTATCGAAAACCTACATGAAATATTTGAAGTC 120  
Qy 791 CCTCAAGTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 850  
Db 121 CCTCAAGTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180  
Qy 851 TGAAGATCAATCCCTAATATGAAATAGGTTATATTTGATTCATGAGGTAAACAGTCAATA 910  
Db 181 TGAAGATCAATCCCTAATATGAAATAGGTTATATTTGATTCATGAGGTAAACAGTCAATA 240  
Qy 911 TAAATTAATGAAAGATGAGATGCTGAAATATCAAGAGAGACAGCCATAGTCTTTACAGT 970  
Db 241 TAAATTAATGAAAGATGAGATGCTGAAATATCAAGAGAGACAGCCATAGTCTTTACAGT 300  
Qy 971 GCTCTGTGGTGTGCTGCTCAAACTGAATGGGTGGGAAAAGGTATGTCATATATAAAG 1030  
Db 301 GCTCTGTGGTGTGCTGCTCAAACTGAATGGGTGGGAAAAGGTATGTCATATATAAAG 360  
Qy 1031 TTCCATTTTGGCCATTATTTGGCAATCTTGCTTTGTTATTTTGGTGGCAGTGTTCCT 1090  
Db 361 TTCCATTTTGGCCATTATTTGGCAATCTTGCTTTGTTATTTTGGTGGCAGTGTTCCT 420  
Qy 1091 GCTTAATCAATTTGCTTTTGTGGCAATCTGCTTTATTTACTGTGACCAAGTCAAGTTTA 1150  
Db 421 GCTTAATCAATTTGCTTTTGTGGCAATCTGCTTTATTTACTGTGACCAAGTCAAGTTTA 480  
Qy 1151 CATCTGTCTTACTCTCTCTCTCCAGGTAAATTCATATATTTGACATCCAGCTAAGA 1210  
Db 481 CATCTGTCTTACTCTCTCTCTCCAGGTAAATTCATATATTTGACATCCAGCTAAGA 540  
Qy 1211 GGGCCCATCT 1270  
Db 541 GGGCCCATCT 600  
Qy 1271 TTAAGTGGGCAATCATTTGTAATGTAATGAAATTAATGTAATGTAATGTAATGTAATG 1330  
Db 601 TTAAGTGGGCAATCATTTGTAATGTAATGAAATTAATGTAATGTAATGTAATGTAATG 660  
Qy 1331 TAAATGCTACTGAGACATCTAGATCAATCAATCAATCAATCAATCAATCAATCAATG 1373  
Db 661 TAAATGCTACTGAGACATCTAGATCAATCAATCAATCAATCAATCAATCAATCAATG 703

RESULT 11  
B0683427/c 710 bp mRNA linear EST 07-OCT-2002  
LOCUS B0683427  
DEFINITION UI-CF-ECL-ace-b-14-0-UI s1 UI-CF-ECL Homo sapiens cDNA clone  
UI-CF-ECL-ace-b-14-0-UI 3', mRNA sequence.  
ACCESSION B0683427  
VERSION B0683427.1 GI:23535356  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 710)  
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.  
TITLES Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
PUBMED 8889548  
COMMENT Contact: McCray, PB  
McCray Lab  
University of Iowa  
2024 University of Iowa Med Lab, Iowa City, IA 52242, USA  
Tel: 319 356 4866

Fax: 319 356 7171  
 Email: paul-mccray@uiowa.edu  
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.reagen.com) or from Open Biosystems  
 (www.openbiosystems.com).  
 The following repetitive elements were found in this CDNA  
 sequence: 579-628, >LINE2  
 Seq primer: M3 FORWARD  
 POLYA=Yes.

#### FEATURES

Location/Qualifiers

1..710  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-CF-ECl-ace-b-14-0-UI"  
 /tissue\_type="lung"  
 /dev\_stage="Adult and Fetal"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_11b="UI-CF-ECl"  
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
 modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 UI-CF-ECl is a normalized cDNA library containing the  
 following tissue(s): Normal lung from adult and from fetal  
 day 64, day 87, week 19 and week 42. The library was  
 constructed according to Bonaldo, Lennon and Soares,  
 Genome Research, 6:791-806, 1996. First strand cDNA  
 synthesis was primed with an oligo-dT primer containing a  
 Not I site. Double stranded cDNA was ligated to an EcoR I  
 into pT73-Pac vector. The oligonucleotide used to prime  
 the synthesis of first-strand cDNA contains a library tag  
 (d)18 tail. The sequence tag for this library is  
 AAGTCTTAC.  
 TAG TSSUS=Normal Lung Epithelial Cells Tissue nos 369-371  
 and 380-383  
 TAG\_LIB=UI-CF-ECl  
 TAG\_SEQ=AAGTCTTAC"

#### ORIGIN

Query Match 34.3%; Score 642; DB 13; Length 710;  
 Best Local Similarity 99.9%; Pred. No. 1,7e-140;  
 Matches 692; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1165 ACTCTCTCCAGGTAATTCATTAATTTGATGATCAGCTAAGAGGCCCATCTTTC 1224  
 710 ACTCTCTCCAGGTAATTCATTAATTTGATGATCAGCTAAGAGGCCCATCTTTC 651

1225 TCACCTCTTCCAGTACATTAATTCAGCAATATTTATTTAGACCCCTTACTGTGGGCAA 1284  
 650 TCACCTCTTCCAGTACATTAATTCAGCAATATTTATTTAGACCCCTTACTGTGGGCAA 591

1285 TCATTTGACGGAATTTGGAATAATGATTAATTCCTTATTCAGTAATGTCTACTGA 1344  
 590 TCATTTGACGGAATTTGGAATAATGATTAATTCCTTATTCAGTAATGTCTACTGA 531

1345 GCACAACTAGTAATCATTTACATGAGGCTCATTTGTTTGAAGGTGTTATTTCA 1404  
 530 GCACAACTAGTAATCATTTACATGAGGCTCATTTGTTTGAAGGTGTTATTTCA 471

1405 TAAACAATTTTACACCAATTCATGATGATTAATTTAGACCAATATATGATCAAGGA 1464  
 470 TAAACAATTTTACACCAATTCATGATGATTAATTTAGACCAATATATGATCAAGGA 411

1465 TAAGTAATTTGAGGTATTCGCAATTTAAAGTATCCGTAATTTGATCACTATTATA 1524  
 410 TAAGTAATTTGAGGTATTCGCAATTTAAAGTATCCGTAATTTGATCACTATTATA 351

1525 AATAATGAAAAAATGATTAATCTGTAATAAAGTGTATTTGTCAGTCACTGTAATAT 1584

Db 350 AATAATGAAAAAATGATTAATCTGTAATAAAGTGTATTTGTCAGTCACTGTAATAT 291

QY 1585 ACTAGAGTATTAATAATTTGTTTACTCTGCTCACCACCAACATGCTAGATTAACCC 1644

Db 290 ACTAGAGTATTAATAATTTGTTTACTCTGCTCACCACCAACATGCTAGATTAACCC 231

QY 1645 CAAAATAGTATTAATCTTGATTAAGTATTAAGAGACTGGTCTATTAATTAAGATTA 1704

Db 230 CAAAATAGTATTAATCTTGATTAAGTATTAAGAGACTGGTCTATTAATTAAGATTA 171

QY 1705 TTTTGAGCAGACAGAGAGCTGTATTCCTAATCTGATTTAGTATGTTCTGTAATGAGAA 1764

Db 170 TTTTGAGCAGACAGAGAGCTGTATTCCTAATCTGATTTAGTATGTTCTGTAATGAGAA 111

QY 1765 ATGTTACCAATTAATTTTACTTTTACTGATTTACATGATCATTTTATGAGGACATGTTCT 1824

Db 110 ATGTTACCAATTAATTTTACTTTTACTGATTTACATGATCATTTTATGAGGACATGTTCT 51

QY 1825 GTGATTAAGCAATTAATCTTTTATGATCA 1857

Db 50 GTGATTAAGCAATTAATCTTTTATGATCA 18

RESULT 12  
 BM921276 1038 bp mRNA linear EST 12-MAR-2002  
 LOCUS BM921276  
 DEFINITION AGNCOURT\_6626202 NIH\_MGC\_115 Homo sapiens CDNA clone IMAGE:5752662  
 VERSION BM921276  
 KEYWORDS BM921276.1 GI:19371655  
 SOURCE EST.  
 ORGANISM Homo sapiens (human)

REFERENCE  
 AUTHORS Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 JOURNAL 1 (bases 1 to 1038)  
 COMMENT NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: c9abbs-remail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LIML at:  
 http://image.llnl.gov  
 Plate: LLAM12787 row: d column: 07  
 High quality sequence stop: 728.

#### FEATURES

source

1..1038  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5752662"  
 /lab\_host="DH10B"  
 /clone\_11b="NIH\_MGC\_115"  
 /note="Organ: pooled brain, lung, testis; Vector:  
 pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA  
 source anonymous pool of 6 male brains, age range 23-27; 1  
 male lung, age 27; and 1 male testis, age 69. Library is  
 oligo-dT primed and directionally cloned (EcoRV site is  
 destroyed upon cloning). Average insert size 1.8 kb,  
 insert size range 1-3 kb. Library is normalized and  
 enriched for full-length clones and was constructed by C.  
 Gruber (Invitrogen). Research Genetics tracking code  
 021. Note: this is a NIH\_MGC Library."

#### ORIGIN

Query Match 34.2%; Score 640; DB 12; Length 1038;  
 Best Local Similarity 99.9%; Pred. No. 3.6e-140;

Matches	760;	Conservative	0;	Mismatches	0;	Indels	1;	Gaps	1;
QY	36	TGAGCCGCGACGCGTTTGTGAGCCCGTTAGTGCGCCCGGCGAGACACCGCCGCCAT	95						
DB	8	TGAGCCGCGACGCGTTTGTGAGCCCGTTAGTGCGCCCGGCGAGACACCGCCGCCAT	67						
QY	96	GTCCCGCTACTGCGTGTCCCGGCGAGACGCTCTTGTGACGAAAGCTGGCGGCGACAC	155						
DB	68	GTCCCGCTACTGCGTGTCCCGGCGAGACGCTCTTGTGACGAAAGCTGGCGGCGACAC	127						
QY	156	CAGGCTGAGAGACTTGCGGCGTGAATTTGCTGATAGTCTGATGATGATGATGAT	215						
DB	128	CAGGCTGAGAGACTTGCGGCGTGAATTTGCTGATAGTCTGATGATGATGATGAT	187						
QY	216	TCCACTGATTTTCACTGCGCGTCCAGAGAGATTTGCTTATTTCAATTTGAGATGT	275						
DB	188	TCCACTGATTTTCACTGCGCGTCCAGAGAGATTTGCTTATTTCAATTTGAGATGT	247						
QY	276	TCTGATGCTGAAGACGCTTTACATTAATTTGAGAGAAAGTGTGGACGCGACAT	335						
DB	248	TCTGATGCTGAAGACGCTTTACATTAATTTGAGAGAAAGTGTGGACGCGACAT	307						
QY	336	TGAATACAGTTTGCGCGGAGATCGAAAGACCAATGAGATGAAGCCAGAGAG	395						
DB	308	TGAATACAGTTTGCGCGGAGATCGAAAGACCAATGAGATGAAGCCAGAGAG	367						
QY	396	GAGGAGTGTGACAGTTTCTTCACTGCTATGATGATTAACAGATACAGATTTCTAGAG	455						
DB	368	GAGGAGTGTGACAGTTTCTTCACTGCTATGATGATTAACAGATACAGATTTCTAGAG	427						
QY	456	CCGAGATTTGAGAGAGAGATCAAGAGTGGCTTTGATTAACAATATAGAGATC	515						
DB	428	CCGAGATTTGAGAGAGAGATCAAGAGTGGCTTTGATTAACAATATAGAGATC	487						
QY	516	GATATGCTCTAGAGACAGTACAGCAGTGAAGACCAAGCGGCTAG-AGAGCCATTCCGA	574						
DB	488	GATATGCTCTAGAGACAGTACAGCAGTGAAGACCAAGCGGCTAG-AGAGCCATTCCGA	547						
QY	575	CATGATAGACCAAACTGAGCTGAGATCCAGATACGATTTCTGTTACTTACTTCAAG	634						
DB	548	CATGATAGACCAAACTGAGCTGAGATCCAGATACGATTTCTGTTACTTACTTCAAG	607						
QY	635	AAAGATCTGAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	694						
DB	608	AAAGATCTGAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	667						
QY	695	GGAAGGTGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	754						
DB	668	GGAAGGTGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	727						
QY	755	GATGTTATCGAAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT	795						
DB	728	GATGTTATCGAAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT	768						
RESULT 13									
LOCUS	BX408797	895 bp	mRNA	linear	EST 13-MAY-2003				
DEFINITION	BX408797 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone								
ACCESSION	BX408797								
VERSION	BX408797.1	GI:30652838							
KEYWORDS	EST.								
SOURCE	Homo sapiens (human)								
ORGANISM	Homo sapiens								
REFERENCE	Bkayvota; Melzosa; Chordata; Craniata; Vertebrata; Euteleostomi;								
AUTHORS	1. (bases 1 to 895)								
TITLE	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.								
JOURNAL	Full-length cDNA libraries and normalization								
COMMENT	Unpublished (2001)								
	Contact: Genoscope								
	Genoscope - Centre National de Sequencage								

QY	798	GTTCGAAAGTGAAGATTTTGAAGCAATTAAGAGAAATTCACCTTGTACTTGGAGAAC	857						
DB	22	GTTCGAAAGTGAAGATTTTGAAGCAATTAAGAGAAATTCACCTTGTACTTGGAGAAC	81						
QY	858	TAAATCCCTAATTAATTAAGTTTATTTGATTCATGAGTAAAGTCCATTAATTA	917						
DB	82	TAAATCCCTAATTAATTAAGTTTATTTGATTCATGAGTAAAGTCCATTAATTA	141						
QY	918	TGGAAGCTAGAGTGTGAATATCAAGAGACAGCCATGCTCTTACAGTGGCTCTG	977						
DB	142	TGGAAGCTAGAGTGTGAATATCAAGAGACAGCCATGCTCTTACAGTGGCTCTG	201						
QY	978	TTGGTCTGTCAAACTGATTTGGGTGGAGAGAGTATGTCATTAATTAAGTCCAT	1037						
DB	202	TTGGTCTGTCAAACTGATTTGGGTGGAGAGAGTATGTCATTAATTAAGTCCAT	261						
QY	1038	TTTGGCATTAATGGCAAACTGCTTGTGTTATTTGGTGGCAAGTCTTCTGCTAAT	1097						
DB	262	TTTGGCATTAATGGCAAACTGCTTGTGTTATTTGGTGGCAAGTCTTCTGCTAAT	321						
QY	1098	CATTGCTTGTGGCATGCTGTTATTTACTTGTACCAAGAGGATTAATCTGT	1157						
DB	322	CATTGCTTGTGGCATGCTGTTATTTACTTGTACCAAGAGGATTAATCTGT	381						
QY	1158	CTTAATCTCTCTTCCAGAGTAAATTCGAATTAATTTGATTAATTCAGTAAAGGCGCA	1217						
DB	382	CTTAATCTCTCTTCCAGAGTAAATTCGAATTAATTTGATTAATTCAGTAAAGGCGCA	441						
QY	1218	TCTCTCTCAAGCTCTTCCAGAGTAAATTCGAATTAATTTGATTAATTCAGTAAAGGCGCA	1277						
DB	442	TCTCTCTCAAGCTCTTCCAGAGTAAATTCGAATTAATTTGATTAATTCAGTAAAGGCGCA	501						
QY	1278	GAGCAATCATTTGATGATTAATTAAGAGAAATTAATTAATTTGATTAATTCAGTAAAGGCGCA	1337						
DB	502	GAGCAATCATTTGATGATTAATTAAGAGAAATTAATTAATTTGATTAATTCAGTAAAGGCGCA	561						
QY	1338	CTAAGAGCAATCTAGTAAATTAATTAAGAGAAATTAATTAATTTGATTAATTCAGTAAAGGCGCA	1397						
DB	562	CTAAGAGCAATCTAGTAAATTAATTAAGAGAAATTAATTAATTTGATTAATTCAGTAAAGGCGCA	621						
QY	1398	TTAATCAATCAATTTTACACCATTCATCATG	1434						
DB	622	TTAATCAATCAATTTTACACCATTCATCATG	658						

## ORIGIN

Query Match 34.0%; Score 637; DB 13; Length 895;  
 Best Local Similarity 100.0%; Pred. No. 2e-139;  
 Matches 637; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## FEATURES

Location/Qualifiers  
 1..895  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DF019YN02"  
 /issue\_type="FETAL BRAIN"  
 /dev\_stage="fetal"  
 /clone\_id="Homo sapiens FETAL BRAIN"  
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

BP 191 91006 EVRY cedex - France  
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 798.r For more information about this cluster, see  
 http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0BAA010ZC02\_CS00876\_1&cluster=798.r.  
 Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0BAA010ZC02\_CS00876\_1.

LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
BU621746/c	BU621746	UI-H-Fli-bga-p-10-0-UI.s1	670 bp mRNA linear EST 23-SEP-2002	UI-H-Fli-bga-p-10-0-UI 3', mRNA sequence.	NCI CGAP Fli Homo sapiens CDNA clone	1 (bases 1 to 670)	NCI-CCGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	Unpublished (1997)	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: James Martin CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, <a href="mailto:bento-soares@iowa.edu">bento-soares@iowa.edu</a> The following repetitive elements were found in this CDNA sequence: 71-120, >LINE2 Seq primer: M13 FORWARD POLY(A)=yes.

FEATURES	Location/Qualifiers
source	1. .670

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-H-Fli1-b9a-p-10-0-UI"  
 /tissue\_type="Cell lines"  
 /dev\_stage="Adult"  
 /lab\_host="MD103 (Life Technologies)"  
 /clone\_lib="NCI CGAP\_Fli1"  
 /note="Organ: Chondrosarcoma; Vector: pT773-Pac  
 (Pharmacia) with a modified polylinker; Site 1: EcoR I;  
 Site 2: Not I; NCI CGAP Fli1 is a normalized cDNA library  
 derived from a pool of mRNA obtained from 4 cell lines  
 from grade III chondrosarcoma tissues. The library was  
 constructed according to Bonaldo, Lemon and Soares,  
 Genome Research, 6:791-806, 1996. First strand cDNA  
 synthesis was primed with an oligo-dT primer containing a  
 Not I site. Double stranded cDNA was ligated to an EcoR I  
 adaptor, digested with Not I, and cloned directionally  
 into pT773-Pac vector. The oligonucleotide used to prime  
 the synthesis of first-strand cDNA contains a library tag  
 sequence that is located between the Not I site and the  
 (dT)18 tail. The sequence tag for this library is  
 GAGGTCGCGTG. The cell lines were provided by Dr. James  
 Martin from the University of Iowa.  
 TAG\_TISSUE=human Chondrosarcoma Grade 3 cell line mix  
 TAG\_LIB=UI-H-Fli1  
 TAG\_SEQ=GAGGTCGCGTG"

Query Match	32.2%	Score 603	DB 131	Length 670
Best Local Similarity	99.8%	Pred. No. 2.1e-131		
Matches 653	Conservative	0	Mismatches 1	Indels 0
				Gaps 0

Qy	697	AAGGCGCTCAATGAATATCTGACCAATATTTTGACTCGGCTCAAAAAAGTATAAGA	756
Db	670	AAGGCGCTCAATGAATATCTGCAATTTTGACTCGGCTCAAAAAAGTATAAGA	611
Qy	757	ATGTTATGCAAACTACATGGAATATATGAAGTCCCTTCAAGTTGAAAGTAAGCATTTT	816
Db	610	ATGTTATGCAAACTACATGGAATATATGAAGTCCCTTCAAGTTGAAAGTAAGCATTTT	551

QY	817	AGGCAATATAAAGGAATTCACCTTGTACTGTGGGAACTAAATCCCTAAATATGADA	876
Db	550	AGCAAAATATAAAGGAATTCACCTTGTACTGTGGGAACTAAATCCCTAAATATGADA	491
QY	877	GGTTTATTTTATTCATGGGTAAACAAGTCCATTAATTAATTTGGAACTAGATGTCTG	936
Db	490	GGTTTATTTTATTCATGGGTAAACAAGTCCATTAATTAATTTGGAACTAGATGTCTG	431
QY	937	AATATCAAAGAGAAGACAGCCATAGTCTCTTAACAGTCCCTGTGTGGTCAGTCAACGTA	996
Db	430	AATATCAAAGAGAAGACAGCCATAGTCTCTTAACAGTCCCTGTGTGGTCAGTCAACGTA	371
QY	997	ATTGGGTGGGAAAAGTATGGTCCAAATAAAGTTCATTTTGGCTATTAATGGCAAT	1055
Db	370	ATTGGGTGGGAAAAGTATGGTCCAAATAAAGTTCATTTTGGCTATTAATGGCAAT	311
QY	1057	CTTGCGTTTGGTTAATTTTGGGCGAAGTCTTTCGCTAAATCTTGGTTGGTGGCATC	1117
Db	310	CTTGCGTTTGGTTAATTTTGGGCGAAGTCTTTCGCTAAATCTTGGTTGGTGGCATC	251
QY	1117	TGTGTTTATTTTACTGTGTACACACATGACGTTTACATCTGCTTAACACTCTCTCCAG	1176
Db	250	TGTGTTTATTTTACTGTGTACACACATGACGTTTACATCTGCTTAACACTCTCTCCAG	191
QY	1177	GTAATTCGCAATTAATTTTGACATCCGAGTAAAGGCGCATCTCTCTGACCTCTTCC	1236
Db	190	GTAATTCGCAATTAATTTTGACATCCGAGTAAAGGCGCATCTCTCTGACCTCTTCC	131
QY	1237	TAGTCACTATATTCAGCAAAATTTTATTTGAGCCCTTACTGTGGGCAATCATTTGACTGG	1296
Db	130	TAGTCACTATATTCAGCAAAATTTTATTTGAGCCCTTACTGTGGGCAATCATTTGACTGG	71
QY	1297	ATAATTGAGAAAAATAGTATATTTCCCTTATTCAGTAATGTCTACTGAGACAA	1350
Db	70	ATAATTGAGAAAAATAGTATATTTCCCTTATTCAGTAATGTCTACTGAGACAA	17

RESULT	15
LOCUS	BM728917
DEFINITION	760 bp mRNA linear EST 01-MAR-2002
ACCESSION	U1-E-B01-aiv-h-08-0-U1.r1 U1-E-B01 Homo sapiens cDNA clone
VERSION	U1-E-B01-aiv-h-08-0-U1.5 , MAFNA sequence.
KEYWORDS	BM728917 GI:19050250
SOURCE	EST.
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens
AUTHORS	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 760) Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
FUNDED	8889548
COMMENT	Contact: Soares, MB

Email: bentlo-soares@iowa.edu  
Tissue Procurement: Dr. Gregg Hageman  
cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research Genetics ([www.resgen.com](http://www.resgen.com)).  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. .760

